LA-UR-15-29210 (Accepted Manuscript)

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

Leibly, David J.
Terwilliger, Thomas Charles
Waldo, Geoffrey S.
Arbing, Mark A.
Pashkov, Inna
DeVore, Natasha
Yeates, Todd O.

To be published in: Structure

DOI to publisher's version: 10.1016/j.str.2015.07.008

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-15-29210

A Suite of Engineered GFP Molecules for Oligomeric Scaffolding

David J. Leibly[1,2], Mark A. Arbing[2], Inna Pashkov[2], Natasha DeVore[3], Geoffrey S. Waldo[3], Thomas C. Terwilliger[3], and Todd O. Yeates[1,2*]

Author Affiliations:

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA [2]UCLA-DOE Institute of Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA

[3]Bioscience Division, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA

*To whom correspondence should be addressed:

Todd O. Yeates

UCLA Department of Chemistry and Biochemistry

611 Charles Young Dr. East

Los Angeles, CA 90095

yeates@mbi.ucla.edu

Keywords: protein scaffolds; protein crystallization; protein engineering; protein assembly; synthetic biology

Abbreviations:

GFP: green fluorescent protein, PDB: Protein Data Bank

PDB Accession Numbers

4W69, 4W6A, 4W6B, 4W6C, 4W6D, 4W6F, 4W6G, 4W6H, 4W6I, 4W6J, 4W6K, 4W6L, 4W6M, 4W6N, 4W6O, 4W6P, 4W6R, 4W6S, 4W6T, 4W6U, 4W72, 4W73, 4W74, 4W7X, 4W75, 4W76, 4W77, 4W7A, 4W7C, 4W7D, 4W7E, 4W7F, 4W7R

SUMMARY

Applications ranging from synthetic biology to protein crystallization could be advanced by facile systems for connecting multiple proteins together in predefined spatial relationships. One approach to this goal is to engineer many distinct assembly forms of a single carrier protein or scaffold, to which other proteins of interest can then be readily attached. In this work we chose green fluorescent protein (GFP) as a scaffold, and engineered many alternate oligomeric forms, driven by either specific disulfide bond formation or metal ion addition. We generated a wide range of spatial arrangements of GFP subunits from 12 different oligomeric variants, and determined their X-ray structures in a total of 33 distinct crystal forms. Some of the oligomeric GFP variants show geometric polymorphism depending on conditions while others show considerable geometric rigidity. Potential future applications of this system are discussed, including its use as a crystallization approach by synthetic symmetrization.

INTRODUCTION

The general idea of connecting and spatially organizing multiple proteins is an emerging theme in synthetic biology. Notable applications include the spatial organization of multiple enzymes for metabolic pathway optimization (Conrado et al., 2008; Dueber et al., 2009; Lee et al., 2012), the organization of signaling molecules (Good et al., 2011; Zeke et al., 2009), and the creation of large self-assembling protein architectures (Lai et al., 2012). Another area under exploration is the synthetic organization of protein molecules into various symmetric forms in order to expand the chances of being able to induce them to form well-ordered crystals (Laganowsky et al., 2011). Facile systems for enabling the specific spatial organization of arbitrary proteins of interest could therefore advance research along various lines.

Ongoing efforts towards engineering proteins for improved crystallization stem from the generally low success rate and unpredictability of macromolecular crystallization (Sundstrom et al, 2006; Stacy et al., 2011). Regardless of the varied explanation for why many proteins are difficult to crystallize, the chances for a successful outcome might be improved by promoting the formation of intermolecular contacts that are compatible with crystal symmetry. Various methods for engineering proteins to improve their likelihood of forming good crystal contacts through surface residue mutations or fusion to a carrier protein, have been described and reviewed (Banatao et al., 2006; Salgado et al., 2008; Forse et al., 2011; Corsini et al., 2008; Moon et al., 2010; Zou & Kobilka, 2012) including fusion to engineered green fluorescent proteins (GFPs) (Suzuki et al., 2010).

Synthetic symmetrization – the engineering of artificially symmetric forms of a given protein molecule – has been promoted as one method for explicitly increasing the likelihood that a protein will be able to form a crystal lattice (Banatao et al., 2006). Two potential advantages have been articulated. First, geometric arguments and analysis of observed crystallization patterns suggests that a modest advantage can be gained by building symmetry into an otherwise asymmetric protein molecule by forcing it to oligomerize. Second and perhaps more important, the ability to produce multiple distinct symmetric forms of a target protein is a major advantage for crystallization. If the protein under study is the subject of crystallization trials, then each of the oligomeric constructs (e.g. specific dimers) is in effect a distinct molecular species with new opportunities to form lattice contacts in the context of a crystal. Distinct dimeric forms of a protein, for example, can be constructed by introducing single cysteine residues at various surface-exposed residues in a protein (Banatao et al., 2006; Forse et al., 2011). In another approach, metal binding half-sites can be designed by introducing two potential metal-ligating residues (e.g. histidines) at proximal positions on the protein surface

(Laganowsky et al., 2011). Those experiments have shown that proteins engineered in such ways form oligomers that are rigid enough for facile crystallization, and that many new opportunities are opened up for the crystallization of a single given protein. In many cases, the new interactions introduced into the target protein contribute to the symmetry of the crystal (Banatao et al., 2006; Chruszcz et al., 2008).

Despite the promise of synthetic symmetrization to expand the opportunities for growing protein crystals, the method as it has been applied so far is experimentally burdensome. Its potential utility is offset by the effort required to repeatedly engineer distinct variants of the target protein. In this study, we explore a route for circumventing that problem. The essential idea is to apply the protein engineering work (i.e. to introduce synthetic symmetrization) to a model protein that can subsequently serve as a general carrier for attaching otherwise arbitrary proteins being targeted for crystallization. As a first choice – though others should be possible – we use GFP as the target for extensive synthetic symmetrization. Prior work has established that GFP can be expressed in split form and then functionally reconstituted from a large fragment and a small fragment (Cabantous et al., 2005 & 2013; Nguyen et al., 2014). Therefore, in principle the large GFP fragment could be engineered to produce many distinct oligomeric forms, and each such oligomeric form would drive the assembly of a target protein that carried the (invariant) small fragment of GFP as a fusion. Such a process separates the engineering efforts (which are performed here on GFP) from the choice of target protein, which only needs to be modified in one way (by fusion to the small fragment of GFP) in order to create multiple distinct forms by complementation. The key elements of the approach are illustrated in Figure 1. The use of monomeric split-GFP to complement and then crystallize another protein bearing a small GFP fragment has been demonstrated already in recent work (Nguyen et al, 2014). Here, the second part of the overall strategy is demonstrated by the construction and crystallographic

investigation of several distinct variants of GFP that were designed to oligomerize in different ways, showing that they are capable of crystallizing in many varied forms. This large suite of engineered GFP proteins thus serves as a foundation for various future developments, including those in the broad area of synthetic biology and in protein crystallization.

RESULTS

Rationale for GFP mediated symmetrization

Engineered 'split' forms of GFP have gained widespread use in the laboratory setting as biosensors (March & Bentley, 2003) or fusion partners to probe for protein solubility (Cabantous et al., 2005 & 2013). These robustly folding mutants of GFP can be expressed without one or more terminal beta-strands of the eleven strands composing the GFP beta barrel. Due to its extensive engineering for stability, the split-GFP is unusually permissive to mutation and topological permutation. Using circular permutants of a full-length GFP containing mutations developed for the split-form of GFP (Cabantous et al. 2005), Bystroff and co-workers created additional split-GFP pairs (with other tagging or "left-out" strands such as beta strand 7) (Huang & Bystroff, 2009). Partial forms of GFP typically lack a mature chromophore (such as GFP missing strand 10 or strands 10 and 11 (Cabantous 2005 & 2013)) or have non-native chromophore environments (as in the circular permutant with strand 7 missing (Huang & Bystroff, 2009)), and likely exist in partially folded states. The partial core can then be complemented by addition of another protein that has been engineered to carry the missing GFP beta strand(s), either as a terminal fusion or as a loop insertion. Once complementation occurs, the full beta barrel is restored and formation of the native chromophore provides a convenient readout of complex formation.

These previous developments make GFP well suited as a general carrier protein for implementing a new approach to the idea of synthetic symmetrization. The particular form of GFP used in our study can be split after strand nine, resulting in the GFP (strands 1-9) core and GFP (strands 10-11) hairpin (Cabantous et al., 2005; Nguyen et al., 2014). With this system, the hairpin formed by strands 10-11 can be engineered into a target protein, which will then complement GFP(1-9). In the simplest scenario, the (10-11) hairpin can be fused as an extension at either the N or C terminus of the target protein. However, the two-stranded hairpin allows for another particularly advantageous kind of construction. If the hairpin can be inserted at an internal sequence position on an exposed loop in the target protein, then the protein complex formed upon complementation will possess two-chain crossing between the reconstituted GFP domain and the target protein structure (Fig. 1). This is expected to enforce a much more rigid spatial arrangement between the two components, which could be an advantage, particularly where crystallization is the ultimate goal. In fact this has been demonstrated in one recent study, where a crystal structure of such a complex revealed two copies of the molecular complex in the asymmetric unit in very nearly the same configuration, suggesting a limited range of motion when using the (10-11) hairpin insertion approach (Nguyen et al., 2014). Anticipating the advantage of the GFP(1-9) plus (10-11) hairpin approach, we focused our efforts in engineering oligomerizing variants of GFP at positions that would be least likely to interfere with subsequent assembly. That is, we primarily engineered regions of GFP remote from the (10-11) hairpin which is ultimately to be carried by the target protein.

Oligomerization strategies

We undertook two approaches to engineering oligomerizing variants of GFP. In the first, individual cysteine residues were introduced at surface positions. Each such

engineered protein was expected to produce a distinctly different dimeric structure upon oxidative disulfide formation. The utility of the disulfide-based approach to synthetic symmetrization has been demonstrated before (Banatao et al., 2006; Forse et al., 2011). The second approach is based on designed metal-mediated interactions following the work of Tezcan *et al.* and Kuhlman *et al.* (Salgado et al., 2008 & 2010, Der et al., 2012). Here, the idea is that introducing a metal half-site into the surface of a protein will lead to assembly upon addition of metal ions (e.g. Ni^{2+}, Zn^{2+}, Cu^{+}). The utility of the metal-mediated approach to synthetic symmetrization has been demonstrated before, where it was found that in addition to the intended dimeric forms; varied modes of assembly can be realized upon metal addition (Laganowsky et al., 2011). In total, in the present work we determine 33 new crystal structures from our series of mutants composed of disulfide-bonded GFP dimers (20 crystal forms), GFP oligomers organized by metal-mediated contacts (seven crystals forms), and cases where disulfide bonds and metal-mediated contacts are both present (six crystal forms) (Tables 1, 2, and S1).

Crystal forms of cysteine dimers

Towards the goal of creating a suite of dimerizing GFP molecules, we created five cysteine point mutations – K26C, D102C, D117C, Q157C and D190C – as well as two sets of mutations to serve as either disulfide or metal-mediated oligomers: E115C/T118H and E124H/K126C. These amino acids were selected for mutation based on their polarity, their surface location, and their distance from strands 10-11 in order to limit interference with complementation when ultimately expressed in the split form (Fig. 2). As the starting or wild type sequence for design of the point mutations, we chose the sequence of Split-GFP in its full-length form (Cabantous et al., 2013) using the superfolder GFP structure as a reference for point mutations in solvent exposed locations (Pedelacq et al., 2006). Two native cysteines at positions C48 and C70 were

mutated to alanine to prevent subsequent interference with disulfide-based dimerization; one exception was an initial experiment and crystal structure of the K26C mutant of the superfolder form (PDB 4W6B) in which only the cysteine at position 48 had been removed.

The ultimate goal of our study is to use engineered versions of the truncated GFP (1-9) to synthetically symmetrize target proteins bearing the (10-11) hairpin, but we judged it prudent to first conduct the GFP engineering experiments in the background of the complete GFP (1-11) construct. Full-length GFP constructs bearing the single engineered cysteine residue were therefore expressed, purified, and then oxidized to form homogenous dimers (Figure 2). For all five of the cysteine sites chosen, pure dimers could be obtained in good yield with ~20-50mg of protein obtained from 2L of auto-induction media.

In order for these engineered GFP dimers to be ultimately useful in crystallizing target proteins using the split protein strategy, we viewed it as a necessary condition that the engineered GFP molecules by themselves must be capable of forming crystals readily. If the engineered dimeric forms of GFP were too flexible to crystallize easily on their own, then they would not be suitable as carrier proteins for crystallizing a target protein in a complex. With the exception of Q175C, crystals grew readily in one to seven days. Depending on the mutant, diffraction quality crystals grew in as few as one condition for K126C or in more than twenty for D102C and D190C.

Due to the large numbers of crystals that grew in the initial experiments, it was not feasible to screen X-ray diffraction in all crystals or to optimize all the crystal hits that were observed. We took the approach of screening crystals that appeared morphologically unique and large enough to mount for X-ray diffraction experiments. In some cases where initial crystals did not diffract despite having good morphology, minor optimization was performed, but otherwise crystals were taken directly from initial

screens. Therefore, it is likely that several additional crystal conditions could have been optimized for various mutants, and that higher resolutions could have been achieved for many of them. Across the many crystal forms examined for the various mutants, the diffraction resolution ranged from 1.7 Å to poorer than 3.5 Å (Table 1). Rather than striving to maximize the resolution for the many crystal forms obtained, we focused on investigating the variety of crystal packing arrangements that these dimers could explore, and the degree to which they appeared to have well-ordered modes of dimerization.

In addition to the cases where we intentionally designed a disulfide bond to make GFP dimers, there were cases where we had anticipated the formation of a metal-binding site between GFP monomers involving a combination of an inserted histidine and cysteine pair, but obtained instead GFP dimers connected by a simple disulfide bond when the metal ion was added. These were mutant pairs D21H/K26C, E115C/T118H and E124H/K126C. In these cases, a disulfide bond was seen in the electron density map, but without evidence for metal binding at the dimer interface. These fortuitous dimers were not explored in depth to try to produce additional crystal forms, so their abilities to form alternative crystal lattices were not established.

In all, we were able to characterize 20 distinctly different crystal forms of the GFP disulfide dimers and solve their structures (Tables 1, S1), with an additional six dimers containing both a disulfide bond and metal contacts. With the exception of the accidental K126C dimer noted above, the various disulfide dimers all crystallized in two or more different space groups. In all these structures, we modeled disulfide bonds into the electron density maps where possible, tabulating standard geometric terms and bond energies for the observed disulfide bonds (Table 2, S2) (Katz & Kossiakoff, 1986). In some cases where the resolution was limited this was not possible, and in at least two cases it appeared that the disulfide bond had been broken during the course of the X-ray

diffraction experiment due to synchrotron radiation damage, as has been observed before (Carugo & Carugo, 2005; Weik et al., 2000).

The occurrence of multiple crystal forms for individual mutants, and the presence in several cases of multiple crystallographically independent GFP dimers in the unit cell, made it possible to analyze the range of conformations and degree of flexibility in these engineered dimers. The disulfide dimers observed and their internal symmetry axes are presented in Figure 3. An analysis of the symmetry and variations due to disulfide bond flexibility was performed for each cysteine mutation by comparing together all dimers that were observed for a given point mutation (Fig. 4, Table 2). In each case we calculated the angle of rotation between the two subunits connected by the engineered disulfide bond to judge whether the synthetically generated dimers were nearly symmetric (i.e. related by a 180° rotation) (Table 2). Then, to evaluate how rigidly connected the two subunits were, we examined the degree of geometric variability between multiple instances of the same dimer as observed across different crystal forms or different asymmetric units of the same crystal form (Tables S3 and S4). For these comparisons one chain of each dimer was designated the A chain and the other B. All the A chains were then aligned and the relative orientations of the B chains in the different instances of the dimer were determined. Particularly for the cases where the dimer was not symmetric, for optimal alignment of multiple dimers it was important to test which of the two chemically identical protein chains should be assigned as the A subunit (the subunit that was superimposed). [N.B. those optimal chain A vs B assignments do not necessarily match those in the deposited PDB files]. A summary of the range of variations for each mutant dimer is presented in Figure 4 and Table 2. A summary of the disulfide-bonded GFP structures is as follows:

K26C – Four crystal forms of K26C dimers were observed (PDB depositions 4W6B, 4W6C, 4W6D and 4W6F), two in each of the space groups $P2_12_12_1$ and $P3_22_11$. Of these, 4W6C was the most symmetric (175.6°) while 4W6F was the least (144.3°). 4W6C, 4W6D and 4W6F were most similar with a maximum variation of 33.3° while 4W6B varied by a rotation of up to 140.4° when overlaid on the others (Table 2, Figure 4B). The dramatically different dimeric arrangement of 4W6B appeared to result from a magnesium ion chelated by Asp19 of each chain helping to stabilize the observed orientations. 4W6D also had a magnesium ion chelated by aspartic acid residues. However with 4W6D it is Asp19 of one chain and Asp21 of the other chelating the ion instead of Asp19 of each chain. One of the structures (4W6F) in which GFP dimers were obtained through a disulfide bond at position 26 arose from a D21H/K26C mutant initially designed for metal chelation. Unexpectedly, addition of Ni^{2+} resulted in formation of a disulfide bond between residues 26C from two protein molecules during the crystallization experiment. The D21H residue of one of the chains and four imidazole molecules from the crystallization condition chelate a nickel ion as well. Structure 4W6C also came from the D21H/K26C mutant, but in this case only a disulfide-bonded dimer was seen, with no metal ions associated with either chain.

D102C – Two crystal forms were observed for this mutant in space groups P1 (4W6R) and $P2_1\,2_1\,2_1$ (4W6P). Crystals appearing in the P1 morphology (thin plates) were obtained in numerous conditions containing PEG polymers as the precipitant, with average molecular weights ranging between 3000-8000 Da. Many of these crystals diffracted only to 7Å or poorer resolution. We were able to solve the structure of 4W6R to 3.47Å; this was the highest resolution we were able to obtain from all the crystals screened of the D102C mutant. This P1 crystal form had a total of eight disulfide-bonded dimers in the crystal asymmetric unit with an average angle between the chains of 167°.

The eight dimers were remarkably similar with a maximum angular variation of only 8° (Figure 4C, Table 2). Due to this small range of variation, the CCP4 program Zanuda (Winn et al., 2011) was used to determine if there was crystallographic symmetry missed in the initial structure determination. This analysis suggested the possibility of the crystal actually being in the C2 space group. However, it was not possible to process the diffraction data satisfactorily in C2. Thus, the true space group of this structure appears to be P1. The 4W6P structure also contained four dimers in the asymmetric unit of P212121. These dimers are less symmetric than those observed in the P1 form (average internal angle between subunits of ~143°). In comparison to the other dimeric forms in the same crystal asymmetric unit of this mutant, one dimer (chains F and G) is a minor outlier, having a relative chain rotation between subunits of 5-8° when compared to the other three dimers. The uniqueness of this dimer effectively rules out the possibility of any higher symmetry in the crystal.

E115C – Originally intended to serve as a metal half-site, the mutated pair of residues, E115C/T118H, revealed disulfide-bonded dimer formation under crystallization conditions with the addition of metal ions. This pair resulted in four structures: three disulfide dimers (4W72, 4W73 and 4W7X) and one structure with metal-mediated contacts only (4W74, discussed subsequently). The three disulfide dimers feature an average rotation angle between subunits of 165°, with a variation up to 12° (Figure 4E, Table 2). Interestingly, 4W72 features a relevant metal-mediated contact as well (Figure 5A); His118 of chain A and Glu17 of chain B chelate a copper ion.

D117C – This mutant resulted in six crystal forms, each in a different space group. The six dimers fall into two groups (Figure 4E, Table 2). Three of the dimeric forms observed (4W6L, 4W6M and 4W6O) are either perfectly symmetric with the two

subunits related by crystal symmetry (4W6L and 4W6O), or very nearly symmetric (4W6M, 179° rotation). 4W6J and 4W6N feature similarly asymmetric dimers (average internal angle of 149°), and 4W6K contains a dimer with an internal angle of 167°. This is an example of a desirable feature in our suite of GFP oligomers. The D117C dimers are rigid enough to form well-ordered crystal lattices, diffracting up to 1.7Å. Yet they are not locked into one conformation, and the permissible angular variation allows for multiple distinct lattices.

K126C – An intended metal-half site pair, E124H/K126C (4W6S) apparently underwent disulfide oxidation in the crystal drop, leading to a symmetric dimer (178°). Copper was added to the protein immediately prior to the crystallization experiment and no copper ions were observed in the crystal structure. No further efforts were undertaken to explore the possibility of additional space groups for this dimer.

Q157C - Two structures were solved from this mutant, 4W69 and 4W6A, and only after screening and optimization of crystal conditions. This is likely a result of the point mutation being located on a somewhat flexible loop of the GFP core. The best crystals diffracted to a resolution of 4 Å (4W69). 4W6A represents an interesting and somewhat mysterious crystal form. Two chains are in the asymmetric unit and they contribute to two different symmetric dimers sitting on axes of crystallographic symmetry, but the expected disulfide bonds are not present. The distance between the cysteine Cα positions of the two subunits is ~11 Å. These crystals took over 6 months to grow, and we suspect that the formate in the crystallization mixture may have slowly reduced the disulfide bonds initially present (Gibson, 1969). Despite the apparent absence of a disulfide bond in the final structure, we calculated intersubunit chain orientations for

comparison (Figure 4F, Table 2). Based on the difficulties crystallizing this mutant, we do not view it as a favorable candidate for future crystallization experiments.

D190C – As with the Q157C point mutation, D190C is located in a flexible loop that is found to be disordered in many of the GFP structures presented in this study. This mutant resulted in >20 conditions with poorly diffracting crystals. We were still able to determine the structures of three D190C mutants (4W6G, 4W6H and 4W6I). 4W6I was the most symmetric dimer (171°) while 4W6G and 4W6H were asymmetric at 141° and 135° respectively (Figure 4G, Table 2).

Taking all the observed disulfide dimers together, we note that only two of these are perfectly symmetric by virtue of lying on crystallographic axes of 2-fold symmetry. Of those that did not fall on symmetry axes, another nine had internal angles between the chains >170° (11 of 36 disulfide dimers observed). The remaining majority of dimers were substantially asymmetric. This contrasts with the trend towards nearly symmetric dimers noted in earlier studies on synthetically symmetrized proteins (Banatao et al., 2006; Forse et al, 2011) that had been connected primarily through alpha helical segments rather than a beta sheet conformation as in GFP.

Metal-mediated oligomer formation

In additional to disulfide dimerization, we explored the possibility of forming dimers or higher oligomers by designing metal binding half-sites in the surface of the GFP molecule. Previous efforts exploring engineered metal-mediation oligomer formation have focused on mutations in alpha helical proteins. In those cases, residues i and i+4 can be mutated to metal-chelating residues (Salgado et al., 2008; Laganowsky et al., 2011). The mutations are typically to His/His or His/Cys pairs, in an attempt to

replicate native chelation motifs. We investigated whether a variation on the approach could be applied to GFP, which consists mainly of a single beta-barrel. We selected residues in three distinct regions of the protein to mutate to either His/His or His/Cys pairs. These mutations were residues i and i+2 on one beta strand (E124/K126) or two residues on adjacent strands (D21/K26 and E115/T118) (Figure 2C). To evaluate their ability to form oligomers in the presence of metal ions, we analyzed purified proteins in the presence of Cu^{2+}, Ni^{2+} and Zn^{2+} salts using native gel shift assays. We determined that mutant pairs D21H/K26C, E115C/T118H, E124H/K126C and E124H/K126H were all able to form oligomers in the presence of each of the ions (Figure 2D). All of these mutant-metal combinations were then used for crystallization experiments to determine their ability to sample different space groups and form metal-mediated crystal contacts. Although D21H/K26H and E115H/T118H did not show shifts on the native gel assay, we proceeded with the crystallization experiments to determine if they could still form metal-mediated contacts during the crystallization process.

From these metal-mediated variants, we solved seven unique structures that were dependent on metal chelation to form. As with the disulfide and mixed disulfide–metal dimers, an ability to crystallize in a variety of conditions was observed with these metal-mediated GFP variants. In a range of other cases, however, the metal ions established crystal contacts between different GFP molecules through a combination of the engineered residues and other native residues (typically Asp and Glu) on the protein surface. Only one of these structures (4W7R) formed a symmetric dimer, whereas the other cases involved more complex spatial arrangements. In several cases, owing to low resolution and poor electron density, it was difficult to determine the exact chelation of the metal ion by the protein side chains. In some instances this likely results from exposure to synchrotron radiation, which can change the oxidation state of metal ions or

damage carboxylic acid groups in the chelating aspartic acid side chains (Carugo & Carugo, 2005; Weik et al., 2000).

D21H/K26C – The designed metal half site mutation D21H/K26C resulted in either disulfide dimers discussed previously or a mixed dimer containing the disulfide and a chelated metal ion (4W75, 4W76, 4W77, 4W7A and 4W7C). In each case the disulfide bond was oxidized during the crystallization experiment. In each structure residues Asp19 and His21 from each chain chelate the copper ion (Figure 5B). Many of these structures have poor electron density for the Asp19 and His21 side chains, and it appears in some instances that only one of the residues from each chain is involved in the metal chelation. Four of the five structures are in the $P2_12_12_1$ space group (4W75, 4W76, 4W77 and 4W7A) but with different packing of the GFP dimers in the unit cell. The fifth structure (4W7C) crystallized in space group C2. Structures 4W76, 4W77, 4W7A and 4W7C are close to being symmetric (average angle of 173.4°) with 4W75 being asymmetric at a 152° inter-subunit rotation. The symmetric structures are very similar to each other, with a variation upon overlap of 2°-8°, while 4W75 differs by up to 32° (Figure 4B, Table 2). As with the pure disulfide dimers, this flexibility can allow the structure to adopt different packing arrangements and crystal forms.

D21H/K26H – Two structures were solved with copper-mediated crystal contacts. The first structure, 4W7E, features one chain in the asymmetric unit with one copper ion present, creating a crystal contact. Here, Asp19 and His21 of one chain and Gln184 of the symmetry mate chelate the ion. This mutant crystallized in the presence of imidazole, leading to one imidazole molecule being associated with the copper ion. Structure 4W7D features two different copper-mediate contacts (Figure 5C) and two chains are present in the asymmetric unit. Chain A makes contacts with two different protein molecules in the

crystal using side chains that were engineered into this mutant. First, His21 and His26 chelate two copper ions and form a crystal contact to Lys3 of one neighboring molecule. A crystal contact to a different molecule is through Lys2 of chain A and Asp19 and His21 of the other protein, similar to the metal chelation observed in the D21H/K26C structures. The high pH (9.5) of this crystallization condition allows the lysine side chain to participate in the chelation of the copper ion.

E115C/T118H – In addition to the observed disulfide dimers of this mutant, structure 4W74 forms a complex system of zinc-mediated crystal contacts between the eight protein chains in the asymmetric unit and six zinc ions via three different coordination sites (Figure 5D). The mutated Cys115/His118 half-site is found to chelate the zinc to a lone Cys115 in two cases; between chain A (Cys115/His118) and chain G (Cy115) and chain D (Cys115/His118) to chain F (Cys115). The Cys115/His118 half site and an aspartic acid residue from a neighboring protein molecule chelate the other four zinc ions in arrangements that are generally similar to each other.

E115H/T118H – Two crystal forms of the E115H/T118H mutant with two different metal-mediated contacts were solved. 4W6U contains four chains in the asymmetric unit, yet only chains A and B feature a nickel-mediated contact. His118 of chain A and His115 of chain B are the residues responsible for metal chelation with a citrate molecule from the crystallization buffer (Figure 5E). A second nickel atom is chelated by residues His25 and Glu132 of chain A alone. In the 4W6T structure, there is one chain in the asymmetric unit that makes contact with other protein molecules through two copper ions (Figure 5F). His115 of the first chain and His25 and Glu132 of the symmetry mate chelate the first copper atom. His118 and Glu32 of the first chain and Asp133 of the symmetry mate chelate the second copper atom.

E124H/K126H – From the final mutant we determined two crystal structures, 4W7F featuring metal-mediated contacts and 4W7R, which is a symmetric metal-mediated dimer. 4W7F contains one chain in the ASU with the copper-mediated contact formed between His124/His126 of the first chain and Glu5 of the symmetry mate (Figure 5G). The only symmetric metal-mediated dimer that we determined the structure of is 4W7R. In this structure the His124/His126 pair of chain A chelates the copper ion with the His124/His126 pair of chain B. Two copper-mediated dimers (four subunits in total) are found in the asymmetric unit, and both dimers are nearly symmetric with chains orientated 179° apart. The two dimers are virtually identical with only a 2° variation when aligned.

GFP oligomers as a crystallization scaffold

After establishing in a previous study that a complex between the split-GFP1-9 and a protein containing the 10/11 hairpin could form diffraction quality crystals (Nguyen et al, 2014), we set out to crystallize a novel protein that had failed to crystallize in previous experiments. We attempted this with the motor domain of STARD9 (Torres et al., 2011), a monomeric kinesin that could serve as a target for novel anti-mitotic drug development. We co-expressed a construct of STARD9 as a N-terminal fusion to the GFP 10/11 hairpin together with our four metal chelating GFP1-9 mutants. We were able to obtain crystals of the STARD9-10/11 and GFP1-9 (D21H/K26C) complex after approximately three months (Fig 6A). However, these crystals are small (~20μM in the largest dimension) and have not produced well-ordered diffraction to date; optimization efforts are underway.

A second computationally designed 271 amino acid protein (to be published) containing the 10/11 hairpin as a loop insertion was co-expressed with the cysteine

mutant suite of split-GFPs. After seven months, triangular plate crystals (~50-75μM) (Fig 6B) were observed containing the designed protein in complex with the GFP1-9 (D117C). As with the STARD9-10/11 constructs, optimization efforts of these crystals are underway.

DISCUSSION

The body of structural data presented here characterizes a suite of engineered GFP molecules comprising a wide range of oligomeric forms, most of which appear highly amenable to crystallization on their own. We obtained 20 new crystal forms of seven disulfide-bonded dimers, plus thirteen metal-mediated structures from five combinations of metal-chelating mutations. Many of the engineered GFPs formed additional crystal forms in numerous conditions that were not pursued for structure determination. The 33 crystal forms are all distinct from each other (Table 1). In analyzing individual GFP variants that were observed in multiple crystal forms, it was found that some of the oligomeric GFPs show strong geometric constraints between the disulfide bonded subunits, while others display considerable geometric polymorphism. Overall the results emphasize the range of geometric arrangements and lattice contacts that can be promoted by the synthetic symmetrization approach.

These engineered GFP proteins provide a system for rapidly creating a series of distinct oligomeric forms of a given target protein, either for crystallization or other applications. Our GFP constructs were engineered to be compatible with use in split form; the oligomerizing mutations are within the main GFP(1-9) fragment and are remote from the (10-11) hairpin, so that engineered variants of the GFP(1-9) construct can be reconstituted with a target protein bearing the (10-11) hairpin. In principle, this reconstitution can be performed in vivo (by co-expression) or in vitro (after separate purifications). Initial experiments with the in vitro approach (not presented here) suggest

that further optimization of the GFP1-9 core may be important in the context of the various mutations introduced into the protein. For those forms based on metal chelation, the identity of the added metal ion provides another convenient variable for modulating the assembly properties of the target protein complex.

A principal long-term motivation for the present work is the crystallization of novel proteins, but other diverse applications in synthetic biology are likely to emerge for these symmetric variants of GFP (Fig 7). One prospective application would be in attaching metabolically coupled enzymes together in different geometries through metal-mediated interactions or *in vitro* oxidized cysteines. Here again, the advantage of the split protein system would be that multiple kinds of configurations could be investigated without having to repeatedly engineer the enzymes under study. They could be used as oligomerizing scaffolds for bringing together homo- or hetero-pairs of proteins into close proximity, in different spatial arrangements, and in ways that can be triggered by the addition of metal ions (Fig. 7 C, D). In order to promote formation of strictly heteromeric assemblies, future experiments would be required to design asymmetric versions of an oligomerizing carrier protein. A final avenue for future applications will be in using oligomerizing carrier proteins (GFP and others that could be developed) to drive other proteins or enzymes to form extended materials or amorphous gels (Fig. 7B). While the motivating application emphasized in the present study (protein crystallization) applies primarily to target proteins that are naturally monomeric, we envision that extended materials, most likely with irregular structures, could be formed by complementing various oligomeric forms of the split-GFP(1-9) with naturally oligomeric proteins or enzymes bearing the 10-11 hairpin. In most cases this would lead to runaway oligomerization. Materials constructed in this way could have novel properties and uses.

METHODS

Cloning

Unless otherwise stated, primers were ordered from Valuegene, enzymes were from New England Biolabs, and DNA sequencing was performed by Genewiz. The plasmid construct containing the split-GFP (Cabantous et al., 2005 & 2013) used as a template to generate a construct with a C-terminal hexahistidine tag and the C-terminus: …TAAGITHHHHHH. The GFP gene was PCR-amplified with Phusion DNA polymerase using the primers GFP.For and GFP.Rev, which include NdeI and HindIII restriction sites, respectively, in the primer extensions. The PCR-amplified segment was purified, digested with NdeI and HindIII and ligated into pET24a, which had been restriction digested with the same two enzymes. Colony PCR using T7 and T7 terminator primers was performed to identify putative positive clones whose DNA sequences were subsequently confirmed by DNA sequencing. Two cysteine residues (Cys48, Cys70) were mutagenized to alanine using the primers C48A.For.New./ C48A.Rev.New. And C70A/ C70A_antisense to eliminate the possibility of unintended disulfide bonds. The C48A mutation was made by linear PCR-amplification of the target vector with Phusion DNA polymerase followed by DpnI digestion of the template plasmid and subsequent phosphorylation of the gel-extracted DNA with T4 polynucleotide kinase and ligation with T4 DNA ligase. The C70A mutation was made using Pfu Turbo AD polymerase (Agilent) using the Quikchange mutagenesis procedure. Additional mutations were made in the GFP construct containing the C48A/C70A mutations by the Quikchange method to generate the following GFP mutant proteins: C48A/C70A/D102C, C48A/C70A/D117C, C48A/C70A/Q157C, C48A/C70A/K26C, C48A/C70A/D190C, C48A/C70A/E124H/K126H and C48A/C70A/E115C/T118H.

Proteins with an N-terminal TEV protease cleavable His6 tag were constructed by cloning the existing GFP mutants in pET24 into a modified pET28 vector with N-

terminal cleavable tag to add the N-terminal sequence: MGSDKIHHHHHHENLYFQG. Briefly, the primers GFP.pMA507-star.For. and GFP.pMA507-star.Rev. were used to PCR-amplify the mutated GFP DNA segments, the DNA was gel extracted, and cloned into pMA507star by the Gibson ISO assembly method(Gibson DG, 2009). pMA507star was PCR-amplified with the primers PIPE.Vec.For. and PIPE.Vec.Rev. to generate compatible DNA overhangs. Primer sequences used are presented in Table S5.

Protein expression

Plasmids containing mutant GFP genes were transformed into BL21-DE3 expression cells (New England Biolabs). 10ml starter cultures were grown with overnight shaking at 37°C in LB media containing appropriate antibiotics. The starter culture was used to inoculate 1L of terrific broth media supplemented with 20ml 50x 5052 auto-induction sugars (Studier, 2005) and appropriate antibiotics. Cultures were grown for 4 hours at 37°C. The temperature was then reduced to 30°C, and cultures were allowed to grow for approximately 20 hours. After growth, the cultures were centrifuged at 5000 x g for 30 minutes at 4°C. Harvested cell paste was stored at -80°C until purification.

Protein purification

Cell paste was thawed at room temperature in a lysis buffer of 20mM Tris pH 8.0, 200mM NaCl, 10mM $MgCl_2$, 30mM Imidazole, 400ug/ml lysozyme, 10ug/ml DNAse and 1mM AEBSF. Once the pellet was thawed, cells were lysed via sonication. Lysed cells were incubated at room temperature for 15 min prior to centrifugation to remove all insoluble material, and lysates were clarified at 25,000 xg for 30 min at 4°C. The soluble lysate fraction was applied to a 5ml Ni-NTA (IMAC) column, rinsed with 10 column volumes of wash buffer consisting of 20mM Tris pH 8.0, 250 mM NaCl, 30mM imidazole. The protein was eluted from the column with wash buffer containing 250mM imidazole.

Elution fractions were pooled and then concentrated until the final volume was approximately 1ml. For the disulfide dimers, the protein was exchanged into a buffer consisting of 20mM Tris pH 9.0, 100mM NaCl. Cysteines where then oxidized to form dimers by the addition of 10ml of dimerization buffer (20mM Tris pH 9.0, 100mM NaCl, 5mM $CuSO_4$). This oxidation reaction was incubated at room temperature for 15 minutes before being quenched by the addition of 50mM EDTA. To separate newly formed dimers from remaining monomers, the protein was dialyzed overnight at 4°C into anion exchange buffer (10mM Tris pH 9.5, 1mM EDTA). The protein was applied to an anion exchange column and then eluted via a salt gradient of 0-1M NaCl in anion exchange buffer. The major peak for each cysteine mutant was assessed for dimer purity by non-reducing SDS-PAGE. Fractions of homogeneous dimers were pooled, buffer exchanged into GFP crystallization buffer (10mM tries, 100mM NaCl), then concentrated to 20mg/ml. Aliquots of protein were flash frozen in liquid nitrogen and stored at -80°C for subsequent crystal trials.

Metal-mediated mutants were purified using the same method, up to the IMAC purification, where the hexahistidine tag was cleaved off with TEV protease overnight at 4°C in TEV cleavage buffer (10mM Tris pH 8.0, 100mM NaCl, 5mM DTT, 1mM EDTA). Cleaved protein was then subject to a second IMAC step to remove the TEV protease, cleaved histidine tag and any uncleaved protein. All unbound protein was pooled, buffer exchanged into crystallization buffer, concentrated to 40mg/ml, flash frozen and stored at -80°C for future crystal trials.

Co-expression with target proteins

The STARD9-10/11 construct consisted of the N-terminal TEV protease cleavable His6 tag (MGSDKIHHHHHHENLYFQG) followed by the 10/11 hairpin sequence, DLPDDHYLSTQTILSKDLNEKRDHMVLLEYVTAAGIT**DAS**, with the 'DAS'

serving as a linker between the hairpin and target protein as previously described (Nguyen, et al., 2014). Only the first 391 amino acids (Met1-Asn391) corresponding to the putative motor domain of the protein were used in this construct.

For the prospective designed protein construct, the GFP 10/11 hairpin was inserted into a presumptive loop between Ser135 and Thr136 of the native 271 amino acid long protein. This construct features a non-cleavable C-terminal His6 tag and as such was not used for the metal mediated experiments.

The expression and purification methods for the co-expressed GFP1-9 and crystallization targets with the 10/11 hairpin were essentially the same as for the GFPs alone. After size exclusion chromatography, the fractions with approximate 1:1 molar ration of GFP1-9 and target protein (visualized by SDS-PAGE) were used for the crystallography experiments.

Crystallization

The GFP oligomers were crystallized using hanging drop vapor diffusion. Initial experiments were carried out at the UCLA crystallization facility using commercial sparse matrix screens in a 96 well format. All initial screening trays were set using a Mosquito liquid handling device (TPP LabTech). Limited optimizations were performed manually in some cases using 24-well Linbro plates. Each disulfide dimer was screened initially with four commercial sparse matrix screens JCSG+ (Qiagen), SaltRx (Hampton Research), Crystal Screen I+II (Hampton Research) and Wizard I+II (EmeraldBio). Metal-mediated mutations were screened with JCSG+ and Wizard only. The final concentration of protein in all crystallization experiments was 20 mg/ml. Metal-mediated mutants were mixed with the metal ions (Ni^{2+}, Zn^{2+}, or Cu^{2+}, in three separate screens) immediately before setting crystal trays, at a final concentration of 20mg/ml protein and 2 mM metal ion salts. Trays were set at room temperature and checked periodically over

30 days. Single crystals were mounted with CrystalCat HT Cryoloops (Hampton Research, Aliso Viejo, CA), cryoprotected as needed. flash frozen with liquid nitrogen, and then screened for diffraction. All diffracting crystals were stored for later data collection. All diffraction data were collected at 100K at APS-NECAT beamline 24-ID-C on a DECTRIS-PILATUS 6M detector. The crystallization and cryoprotectant conditions are reported in Table S6.

Structure determination

Data sets from individual crystals were indexed, integrated and scaled using XDS/XSCALE (Kabsch, 2010), with the resolution limit selected to balance completeness, calculated I/σ, R_{sym} and CC1/2 of the highest resolution shell with emphasis on I/σvalues of >1.5 and CC1/2 values >0.9, this was subject to change depending on the quality of the diffraction data. Structures were solved by molecular replacement using the program Phaser (McCoy et al., 2007), with the superfolder GFP (Pédelacq et al., 2006) protein (PDB 2B3P) as the search model. To accelerate the model building and refinement, molecular replacement solutions were initially refined with the PDB_REDO server (Joosten et al., 2011). Final iterative rounds of model building and refinement were carried out using Coot (Emsley et al., 2010), PHENIX (Adams et al., 2010) with TLS refinement (Painter et al., 2006). Structures were validated with PROCHECK (Laskowski et al., 1993), ERRAT (Colovos & Yeates, 1993), MolProbity (Davis et al., 2007) and VERIFY3D (Luthy et al., 1992). Atomic coordinates and structure factors for all 33 structures were deposited in the PDB with codes; 4W69, 4W6A, 4W6B, 4W6C, 4W6D, 4W6F, 4W6G, 4W6H, 4W6I, 4W6J, 4W6K, 4W6L, 4W6M, 4W6N, 4W6O, 4W6P, 4W6R, 4W6S, 4W6T, 4W6U, 4W72, 4W73, 4W74, 4W7X, 4W75, 4W76, 4W77, 4W7A, 4W7C, 4W7D, 4W7E, 4W7F and 4W7R. The internal axes of

symmetry as depicted in Figure 3 was determined with the program SymD (Kim et al., 2010). Figures depicted the structures were made with PyMOL (Schrödinger, LLC).

Structure comparison procedure

To compare multiple observed instances of the same disulfide-bonded dimer, one structure was first chosen as the reference. Then one chain of a subsequent dimer was aligned to chain A of the reference dimer, and the transformation required for overlapping those two chains was applied to the second chain. Both possible assignments to chain A vs B were tested for each dimer, and the best match was retained for comparison.

Acknowledgements

This work was supported by NIH grant P01 GM098177 (to TCT). DJL was supported by Ruth L. Kirschstein National Research Service Award T32GM007185. The authors thank Michael Sawaya, Duilio Cascio and Michael Thompson for X-ray data collection at APS beam line 24-ID-C. We thank Michael Collazo for help with the crystallization trials and Dan McNamara for help with structure determinations. We thank David Baker and Fabio Parmeggiani for providing the designed protein as a target for fusion-based crystallization experiments. We thank the staff of the NE-CAT synchrotron beamline, including Jon Schuermann, Igor Kourinov, and Malcolm Capel, and for helpful discussions. X-ray data collection was supported by DOE Grant DE-FC02-02ER63421 and the NE-CAT beamlines of the Advanced Photon Source, which are supported by National Institutes of Health Grant RR-15301(NCRR). Use of the Advanced Photon Source is supported by the DOE, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357.

References

Adams P.D., Afonine P.V., Bunkóczi G., Chen V.B., Davis I.W., Echols N., Headd J.J., Hung L.W., Kapral G.J,. Grosse-Kunstleve R.W., McCoy A.J., Moriarty N.W., Oeffner R., Read R.J., Richardson D.C., Richardson J.S., Terwilliger T.C., and Zwart P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Banatao D.R., Cascio, D., Crowley C.S., Fleissner M.R., Tienson H.L., and Yeates T.O. (2006). An approach to crystallizing proteins by synthetic symmetrization. Proc Natl Acad Sci *103*, 16230-16235.

Cabantous S., Terwilliger T.C., and Waldo GS. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nature Biotechnology *23*, 102-107.

Cabantous S., Nguyen H.B., Pedelacq J.D., Koraïchi F., Chaudhary A., Ganguly K., Lockard M.A., Favre G., Terwilliger T.C., and Waldo G.S. (2013). A new protein-protein interaction sensor based on tripartite split-GFP association. Sci Rep *3*:2854.

Carugo O., and Carugo, K.D. (2005). When X-rays modify the protein structure: radiation damage at work. Trends Biochem Sci *30*, 213-219.

Chruszcz M., Potrzebowski W., Zimmerman M.D., Grabowski M., Zheng H., Lasota P., and Minor W. (2008). Analysis of solvent content and oligomeric states in protein crystals—does symmetry matter? Protein Sci *17*, 623–632.

Colovos C.Y. & Yeates T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci *2*, 1511-1519.

Corsini L., Hothorn M., Scheffzek K., Sattler M., and Stier G. (2008). Thioredoxin as a fusion tag for carrier-driven crystallization. Protein Sci *17*, 2070–2079.

Conrado R.J., Varner J.D., and DeLisa M.P. (2008). Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol *19*, 492-499.

Davis I.W., Leaver-Fay A., Chen V.B., Block J.N., Kapral G.J., Wang X., Murray L.W., Arendall W.B. 3rd, Snoeyink J., Richardson J.S., and Richardson D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acids Res *35*, W375-W383.

Der B.S., Machius M., Miley M.J., Mills J.L., Szyperski T., and Kuhlman B. (2012). Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. J Am Chem Soc *134*, 375-385.

Dueber J.E., Wu G.C., Malmirchegini G.R., Moon T.S., Petzold C.J., Ullal A.V., Prather K.L., and Keasling J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nature biotechnology *27*, 753-759.

Emsley P., Lohkamp B., Scott W.G. and Cowtan K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Forse G.J., Ram, N., Banatao D.R., Cascio D., Sawaya M.R., Klock H.E., Lesley S.A., and Yeates T.O. (2011). Synthetic symmetrization in the crystallization and structure determination of CelA from Thermotoga maritima. Protein Sci *20*, 168-178.

Gibson DG, Y.L., Chuang R.Y., Venter J.C., Hutchison C.A., and Smith H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Gibson H.W. (1969). Chemistry of formic acid and its simple derivatives. Chem Rev *69*, 673–692.

Good M.C., Zalatan J.G., and Lim W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. Science *332*, 680-686.

Huang Y.M., and Bystroff C. (2009). Complementation and reconstitution of fluorescence from circularly permuted and truncated green fluorescent protein. Biochemistry *48*, 929-940.

Joosten R.P., Joonsten K., Cohen S.X., Vriend G., and Perrakis, A. (2011). Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank. Bioinformatics *27*, 3392-3398.

Kabsch W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.
Kim, C., Basner, J., and Lee, B. (2010). Detecting internally symmetric protein structures. BMC Bioinformatics *11*, 303.

Katz, B.A. and Kossiakoff A. (1986). The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. J Biol Chem *261,* 15480-5.

Laganowsky A, Zhao M., Soriaga AB, Sawaya MR, Cascio D, and Yeates TO. (2011). An approach to crystallizing proteins by metal-mediated synthetic symmetrization. Protein Sci *20*, 1876-1890.

Lai Y.T., King N.P., and Yeates T.O. (2012). Principles for designing ordered protein assemblies. Trends Cell Biol *22*, 653-661.

Laskowski R.A., Macarthur M. W. Moss, D. S., and Thornton J. M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr *26*, 283-291.

Lee H., DeLoache W.C., and Dueber J.E. (2012). Spatial organization of enzymes for metabolic engineering. Metab Eng *14*, 242-251.

Luthy R., Bowie J. U. and Eisenberg D. (1992). Assessment of protein models with three-dimensional profiles. Nature *356*, 83-85.

March J.C., Rao G., and Bentley W.E. (2003). Biotechnological applications of green fluorescent protein. Appl Microbiol Biotechnol *62*, 303-315.

McCoy A.J., Grosse-Kunstleve R.W., Adams P.D., Winn M.D., Storoni L.C., and Read R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Moon A.F., Mueller G.A., Zhong X., and Pedersen L.C. (2010). A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction. Protein Sci *19*, 901-913.

Nguyen, H.B., Hung L.W., Yeates T.O., Terwilliger T.C., and Waldo G.S. (2014). Split green fluorescent protein as a modular binding partner for protein crystallization. Acta Crystallogr D Biol Crystallogr *69*, 2513–2523.

Painter J.M and Merritt E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr *62*, 439-450.

Pedelacq J.D., Tran T., Terwilliger T.C., and Waldo G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol *24*, 79-88.

Salgado E.N., Ambroggio X.I, Brodin J.D., Lewis R.A., Kuhlman B., and Tezcan F.A. (2008). Metal templated design of protein interfaces. Proc Natl Acad Sci *107*,*1827-32.*

Salgado E.N, Lewis R.A., Faraone-Mennella J., and Tezcan F.A. (2008). Metal-Mediated Self-Assembly of Protein Superstructures: Influence of Secondary Interactions on Protein Oligomerization and Aggregation. J Am Chem Soc *130*, 6082–6084.

Stacy R., Begley D.W., Phan I., Staker B.L., Van Voorhis W.C., Varani G., Buchko G.W., Stewart L.J., and Myler P.J. (2011). Structural genomics of infectious disease drug targets: the SSGCID. Acta Crystallogr Sect F Struct Biol Cryst Commun *67*, 979-984.

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Prot Expres Pur *41*, 207–234.

Sundstrom M., Norin M., and Edwards A. Structural Genomics and High Throughput Structural Biology, 2006. (Boca Raton, FL, USA: CRC Press).

Suzuki N., Hiraki M., Yamada Y., Matsugaki N., Igarashi N., Kato R., Dikic I., Drew D., Iwata S., Wakatsuki S., and Kawasaki M. (2010). Crystallization of small proteins assisted by green fluorescent protein. Acta Crystallogr D Biol Crystallogr *66*, 1059-1066.

Torres J.Z., Summers M.K., Peterson D., Brauer M.J., Lee J., Senese S., GholkarA.A., Lo Y.C., Lei X., Jung, K. Anderson D.C., Davis D.P., Belmont L., and Jackson P.K. (2011). The STARD9/Kif16a kinesin associates with mitotic microtubules and regulates spindle pole assembly. Cell *147*, 1309-1323.

Weik M., Ravelli R.B., Kryger G., McSweeney S., Raves M.L., Harel M., Gros P., Silman I., Kroon J., and Sussman J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc Natl Acad Sci *97*, 623-628.

Winn M.D., Ballard C.C., Cowtan K.D., Dodson E.J., Emsley P., Evans P.R., Keegan R.M., Krissinel E.B., Leslie A.G., McCoy A., McNicholas S.J., Murshudov G.N., Pannu N.S., Potterton E.A., Powell H.R., Read R.J., Vagin A., and Wilson K.S. (2011).

Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Zou Y., Weiss W.I., and Kobilka B.K. (2012). N-Terminal T4 Lysozyme Fusion Facilitates Crystallization of a G Protein Coupled Receptor. PloS One *7*, e46039.

Zeke A., Lukac, M., Lim W.A., and Remenyi A. (2009). Scaffolds: interaction platforms for cellular signalling circuits. Trends Cell Biol *19*, 364-374.

Figure Legends

Figure 1 Concept of split-GFP-mediated synthetic symmetrization. A) Split-GFP serves as a scaffold to induce synthetic symmetry. GFP(1-9) (green) can be expressed separately from GFP(10-11) (red), point mutations (yellow) can then be introduced to the GFP(1-9) core creating a symmetric dimer of GFP. B) The GFP(10-11) hairpin can then be inserted into a permissive loop, or fused terminally, to a target protein of interest, which can be expressed separately from the GFP(1-9) core. C) When mixed *in vitro or expressed together in vivo*, the GFP(1-9) cores complement with the target protein containing GFP(10-11). This can be performed with a series of pre-formed GFP(1-9) dimers, resulting in multiple unique dimers of the split-GFP-target protein complex, each with the ability to explore different possible crystal contacts.

Figure 2 Locations of point mutations introduced on full-length split-GFP to induce oligomerization. A) Locations of the individual point mutations to cysteines on the GFP(1-9) core (green) on the opposite face of the beta-barrel from the GFP(10-11) hairpin (red). B) Each cysteine point mutant was purified in non-reducing conditions and dimer formation was visualized on a non-reducing SDS-PAGE gel. After an initial IMAC step, GFP variants were dimerized with Cu^{2+}. The dimeric form (D) was then separated from the monomer (M) via anion exchange chromatography and used for crystallization experiments. C) Locations of the metal-half site mutations on GFP; each site involves a pair of spatially proximal mutations (indicated). D) Native PAGE screening of each metal chelating mutation in the presence of Cu^{2+}, Ni^{2+} and Zn^{2+}. This screen showed apparent oligomer formation for the D21H/K26C, E115C/T118H, E124H/K126H and E124H/K126C variants, as determined by a mobility shift from the monomeric (M) band to the assumed oligomeric (O) band.

Figure 3 Examples of the GFP dimer observed. The internal rotation axis relating the subunits of each dimer is shown (red dot for disulfide dimers, blue for the mixed dimer and orange for the metal-mediate dimer). For each dimer the rotation axis corresponds to the location of the engineered disulfide bond, or metal-mediated crystal contact. The 12 dimers shown are from structures: A) 4W6B B) 4W6C C) 4W7C D) 4W6R E) 4W7X F) 4W6M G) 4W6G H) 4W6I I) 4W6S J) 4W69 K) 4W6K L) 4W7R. They are representative of the complete set of 43 total dimers visualized in this work.

Figure 3. Chain angle ranges for dimers. Depicted is the range of variation between the chain orientations for each disulfide-bonded dimer. Chain A of each dimer was first aligned to visualize the difference in the orientation of the distinct versions of chain B. Only the chain B backbone traces are depicted. Each panel illustrates the multiple conformations observed for one specific cysteine mutant. The blue and red traces represent the range of orientations the chains adopted. When a single outlier is found it is shown in cyan. When two disparate groups of conformations are present, they are shown in red and blue, and cyan and magenta. When more than one dimer was observed in the asymmetric unit, instances representing the extremes in conformation were chosen. The structures and dimer chains displayed are: A) K26C red: 4W6C, blue: 4W6F, cyan: 4W6B. B) D21H/K26C; red: 4W7A AB dimer, blue: 4W7A CD dimer, cyan: 4W75. C) D102C; red: 4W6P CD dimer, blue: 4W6P FG dimer, cyan: 4W6R AN dimer, magenta: 4W6R KL dimer. D) E115C; red: 4W72, blue: 4W73. E) D117C; red: 4W6O, blue: 4W6K, cyan: 4W6N BF dimer, magenta: 4W6J. F) Q157C; red: 4W69, blue: 4W6A A dimers, cyan: 4W6A B dimer. G) D190C; red: 4W6H, blue: 4W6I, cyan: 4W6G.

Figure 5. Observed metal-mediated crystal contacts

A) Structure 4W72: in addition to a disulfide bond between Cys115 of chain A and B, a copper ion is chelated by His118 of chain A and Glu17 of chain B. B) Example of a mixed dimer from structure 4W76. Here His21 chelates the copper ion and Asp19 of both chain A (blue) and chain B (yellow), forming the dimer in addition to the Cys26-Cys26 disulfide bond. Each instance of the mixed D21H/K26C dimer has some combination of Asp12 and His21 chelating the metal ion. C) The two forms of metal-mediated contacts in 4W7D between residues of chain A (blue) and chains B in adjacent asymmetric units (orange [left] and green [right]). His21 and His26 creating the crystal contact with Lys3 of chain B chelate the two copper atoms (left). Lys3 of chain A then makes a crystal contact via copper chelated to His21 and Asp19 of Chain B (right). D) The three observed zinc-mediated contacts found in 4W74. Cys115/His118 chelate the ion with Cys115 of the other chain (left). Cys115/His118 were also found to chelate the zinc ion with Asp190 (middle) or Asp102 (right) as well. E) Nickel mediated crystal contact of 4W6U, His118 of chain A serves as a half site with His115 of chain B, a citrate molecule is also found chelating the nickel ion. F) A double copper-mediated contact of 4W6T creating the contact between adjacent asymmetric units. His115 Glu32 (which has two conformations) to Asp133 of the adjacent chain chelates the first atom. His118 chelates the second copper atom from the first chain and His25/Glu132 of the second. G) The copper-mediated contact between the asymmetric units of 4W7F. His124 and His126 chelate the copper ion with Glu5 of the symmetry mate. H) Copper-chelation by His124 and His126 of the symmetric dimer of 4W7R.

Figure 6 Crystals of split-GFP with a novel crystallization target. A) Crystals of the STARD9-10/11 – GFP1-9 (D21H/K26C) complex were obtained in a condition composed of 10%v/v 2-Propanol, 0.1M MES pH 6.0 and 0.2M $Ca(OAc)_2$. The protein

complex was mixed in a 1:1 molar ratio with $CuSO_4$ immediately prior to the crystallization experiments. The green color of the crystals is used as an indication of the complex formation; the largest crystals observed to date (~20μM in the largest dimension) are highlighted by the red circle. B) Crystals of a designed protein with an internal 10/11 hairpin in complex GFP1-9 (D117C). The triangular plate crystals (~50-75μM) grew in a condition containing 0.1M SPG buffer pH 5.0 and 25% w/v PEG-1500.

Figure 7 Alternative applications for the oligomeric GFPs. In addition to the utility of a suite of oligomeric GFPs for inducing symmetry and allowing the crystallization of novel proteins, we envision that these engineered proteins will have additional valuable applications. A) Attachment of the dimers may be used to change the crystal forms of existing proteins. Here a disordered crystal (top) can form a different and possibly better-ordered lattice (bottom) through fusion to one of the GFP oligomers in the available suite. B) Fusion to a multimeric enzyme, in this example a tetramer, could be used to create an enzymatically active amorphous gel for facile separation of enzymes and products for *in vitro* reaction systems. C) With the split form or through terminal fusions, the GFP dimers can be used to create a heterodimer for co-localization of enzymes for substrate channeling or crystallization experiments. D) Expanding on the idea from (C), two proteins can be forced into close proximity and further symmetrized, by separate genetic fusion of strand 10 to one protein and strand 11 to the other, then allowing them to complement for various applications.

Table 1 – Summary of New GFP Crystal Forms

PDB	Mutation	Type	Space Group	Resolution (Å)	ASU[#]
4W69	Q157C	Disulfide	P 43 21 2	3.98	2
4W6A	Q157C	Disulfide	P 32 2 1	2.99	2
4W6B	K26C*	Disulfide	P 21 21 21	1.90	2
4W6C	D21H/K26C^	Disulfide	P 21 21 21	2.49	2
4W6D	K26C	Disulfide	P 32 2 1	3.45	2
4W6F	D21H/K26C	Disulfide	P 32 2 1	2.70	2
4W6G	D190C	Disulfide	P 61	3.02	2
4W6H	D190C	Disulfide	P 65	1.95	2
4W6I	D190C	Disulfide	P 21 21 21	2.63	2
4W6J	D117C	Disulfide	P 31 2 1	1.70	2
4W6K	D117C	Disulfide	P 41 21 2	2.88	2
4W6L	D117C	Disulfide	I 41 2 2	2.45	1
4W6M	D117C	Disulfide	P 63	2.79	4
4W6N	D117C	Disulfide	C 1 2 1	3.38	6
4W6O	D117C	Disulfide	P 64 2 2	2.60	1
4W6P	D102C	Disulfide	P 21 21 21	3.09	8
4W6R	D102C^	Disulfide	P 1	3.47	16
4W6S	D124H/K126C	Disulfide	P 43 21 2	3.10	2
4W6T	E115H/T118H	Cu Mediated Contacts	P 43 21 2	1.60	1
4W6U	E115H/T118H	Ni Mediated Contacts	P 21 21 21	2.28	4
4W72	E115C/T118H	Disulfide + Metal Contacts	P 21 21 21	1.99	2
4W73	E115C/T118H	Disulfide	P 21 21 21	2.79	2
4W74	E115C/T118H	Zn Crystal Contacts	P 1 21 1	2.10	8
4W7X	E115C/T118H	Disulfide	P 1 21 1	2.80	4
4W75	D21H/K26C^	Disulfide + Metal Contacts	P 21 21 21	3.47	2
4W76	D21H/K26C^	Disulfide + Metal Contacts	P 21 21 21	2.35	2
4W77	D21H/K26C^	Disulfide + Metal Contacts	P 21 21 21	3.10	2
4W7A	D21H/K26C^	Disulfide + Metal Contacts	P 21 21 21	3.60	4
4W7C	D21H/K26C^	Disulfide + Metal Contacts	C 1 2 1	2.50	4
4W7D	D21H/K26H	Cu Crystal Contacts	P 21 21 21	1.80	2
4W7E	D21H/K26H	Cu Crystal Contacts	P 41 21 2	2.59	1
4W7F	D124H/K126H	Cu Crystal Contacts	C 2 2 21	2.90	1
4W7R	D124H/K126H	Cu Dimers	P 1 21 1	1.80	4

* Superfolder GFP C48A backbone mutation,

^ Split-GFP C48A backbone mutation. All other sequences have the double mutations of C48A and C70A.

[#] Number of GFP chains in the asymmetric unit

Table 2 GFP Disulfide Dimer Characterizations

Mutant	PDB	Dimer	Disulfide Cα Distance (Å)	Dimer Angle (°)	Grouped PDBs	Chain "B" Variation Range (°)
K26C	4W6B	AB	6.4	151.66	Group: 4W6C, 4W6D, 4W6F Outlier: 4W6B	Group 4W6C – 4W6F = 33.3 Maximum Range: 4W6B – 4W6D = 140.4
	4W6C	AB	6.2	175.55		
	4W6D	AB	6.2	158.12		
	4W6F	AB	5.6	144.29		
D21H/K26C	4W7A	AB	5.8	169.72	Group: 4W7A, 4W7C, 4W76 4W77 Outlier: 4W75	Group: 4W7A AB – 4W7A CD = 6.3 Maximum Range: 4W7A CD – 4W75 = 32.1
	4W7A	CD	6.2	177.95		
	4W7C	AB	5.9	173.38		
	4W7C	CD	6.4	171.85		
	4W75	AB	6.2	151.90		
	4W76	AB	6.4	174.64		
	4W77	AV	6.1	173.00		
D102C	4W6P	AB	4.5	143.38	Group 1: 4W6P Group 2: 4W6R	Group 1: 4W6P CD - 4W6P FG = 8.3 Group 2: 4W6R AN - 4W6R KL = 7.7 Maximum Range: 4W6P FG - 4W6R KL = 32.4
	4W6P	CD	4.6	146.21		
	4W6P	EH	4.6	143.79		
	4W6P	FG	4.6	139.64		
	4W6R	AN	5.2	165.37		
	4W6R	BI	4.7	165.15		
	4W6R	CD	4.1	170.66		
	4W6R	EJ	4.4	167.73		
	4W6R	FO	4.7	166.16		
	4W6R	GO	4.9	163.96		
	4W6R	HM	4.9	166.20		
	4W6R	KL	4.3	170.91		
E115C	4W7X	AB	6.2	166.40	Group: 4W7X, 4W72, 4W73	4W72 - 4W73 = 12.3
	4W7X	CD	5.4	163.93		
	4W72	AB	5.9	159.85		
	4W73	AB	6.4	170.95		
D117C	4W6J	AB	5.7	154.89	Group1: 4W6K, 4W6L, 4W6M, 4W6O Group2: 4W6J, 4W6N	Group 1: 4W6O - 4W6K = 16.4 Group 2: 4W6N BF - 4W6J = 10.8 Maximum Range: 4W6N BF - 4W6M AC = 34.8
	4W6K	AB	5.7	166.82		
	4W6L	AB	5.5	180.0		
	4W6M	AC	5.6	178.44		
	4W6M	BD	6.5	178.14		
	4W6N	AD	6.1	148.41		
	4W6N	BF	6.3	146.59		
	4W6N	CE	6.4	146.87		
	4W6O	AB	5.5	179.97		
K126C	4W6S	AB	6.00	177.96	--	--
K126H	4W7R	AB	--	179.1	--	AB – CD = 1.7
	4W7R	CD		179.15		
Q157C	4W69	AB	5.5	141.18	--	4W96 – 4W6A B = 129
	4W6A	A	5.78	180.0		
	4W6A	B	*11.7	180.0		
D190C	4W6G	AB	5.8	140.95	Group: 4W6G, 4W6H Outlier: 4W6I	Group: 4W6G - 4W6H = 6.3 Maximum Range: 4W6H - 4W6I = 41.4
	4W6H	AB	5.8	135.23		
	4W6I	AB	6.4	171.21		

*Potential disulfide broken during crystallization.

Figure 1

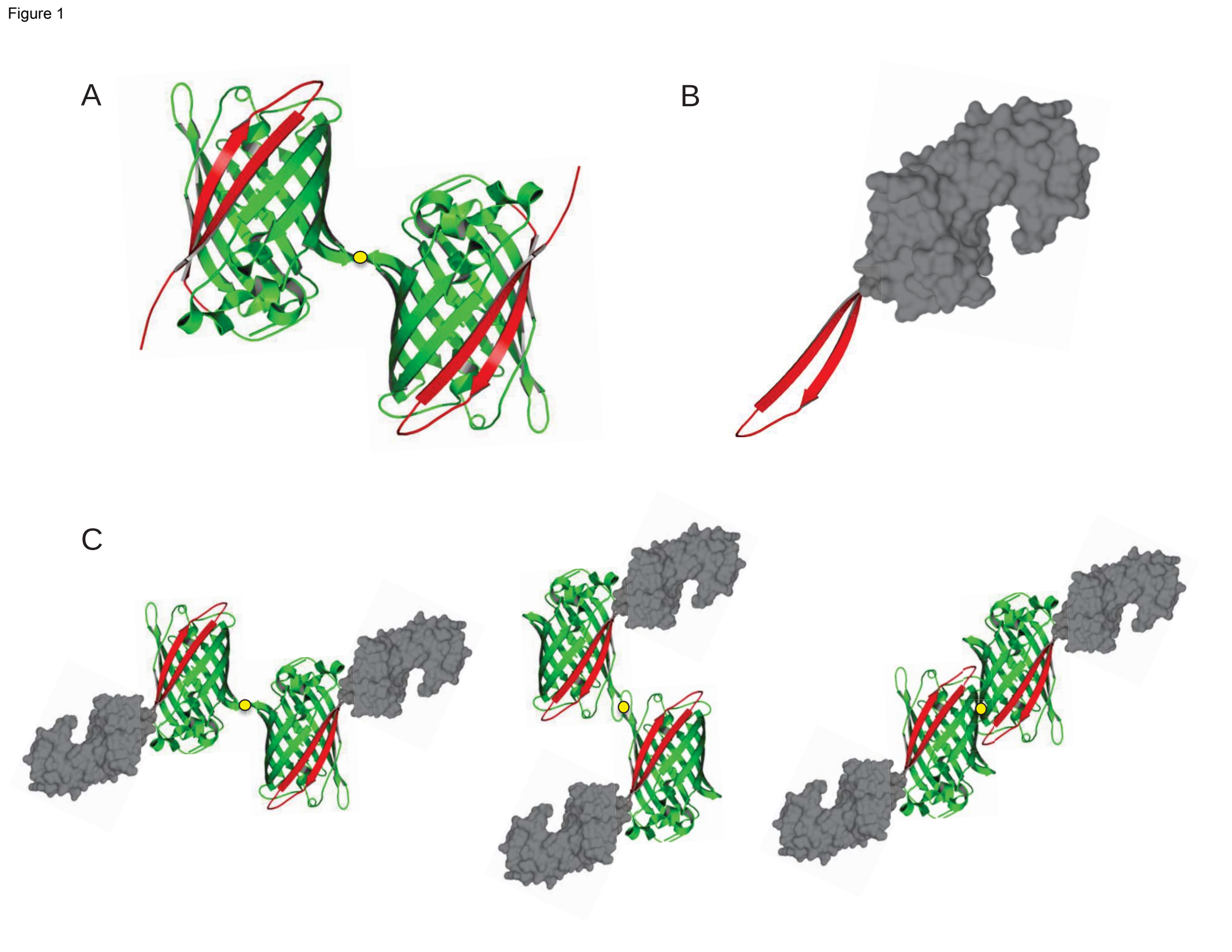


Figure 2

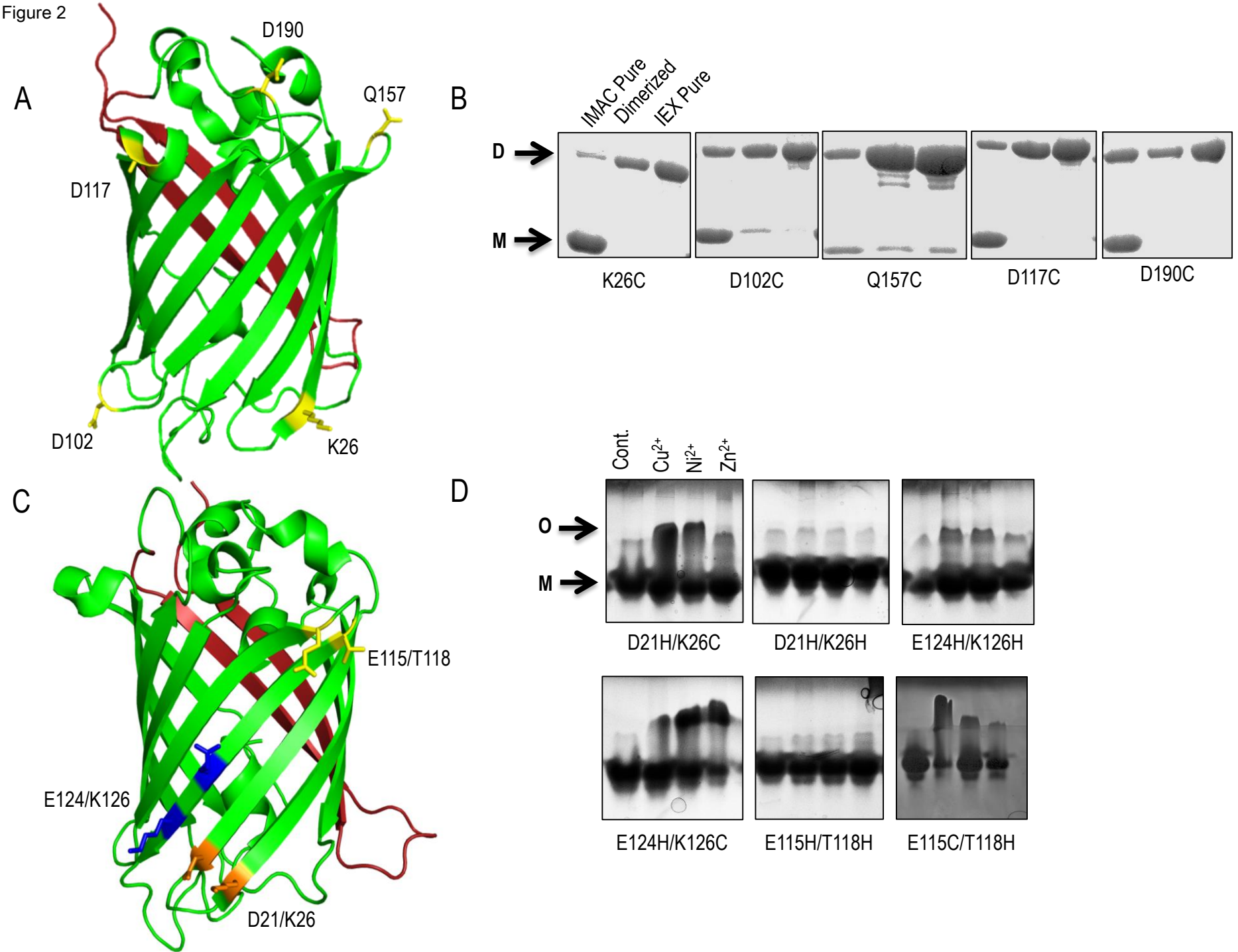

A
D190
Q157
D117
D102
K26
B
IMAC Pure
Dimerized
IEX Pure
D
M
K26C
D102C
Q157C
D117C
D190C
C
E115/T118
E124/K126
D21/K26
D
Cont.
Cu2+
Ni2+
Zn2+
O
M
D21H/K26C
D21H/K26H
E124H/K126H
E124H/K126C
E115H/T118H
E115C/T118H

Figure 3

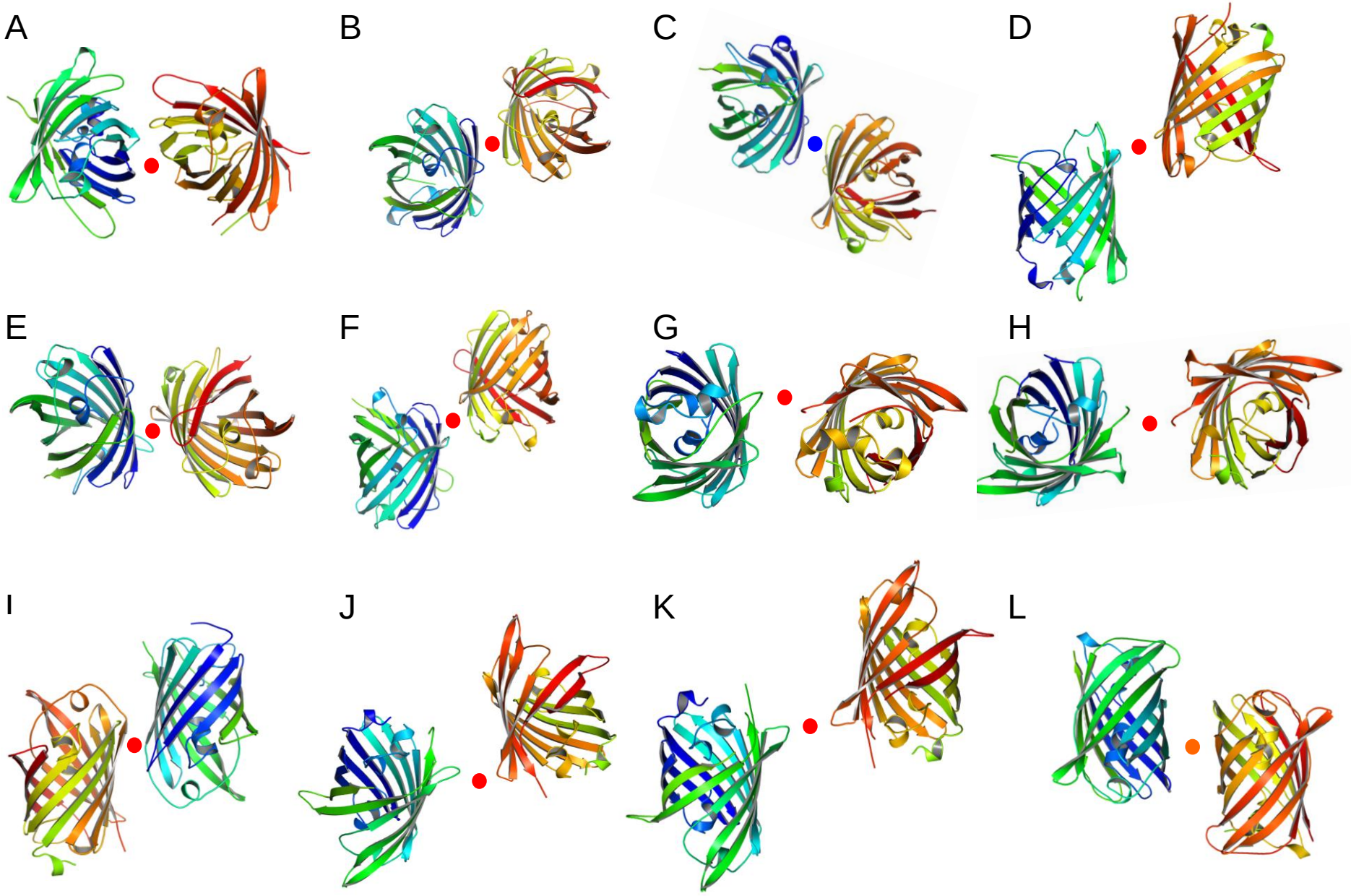


Figure 4

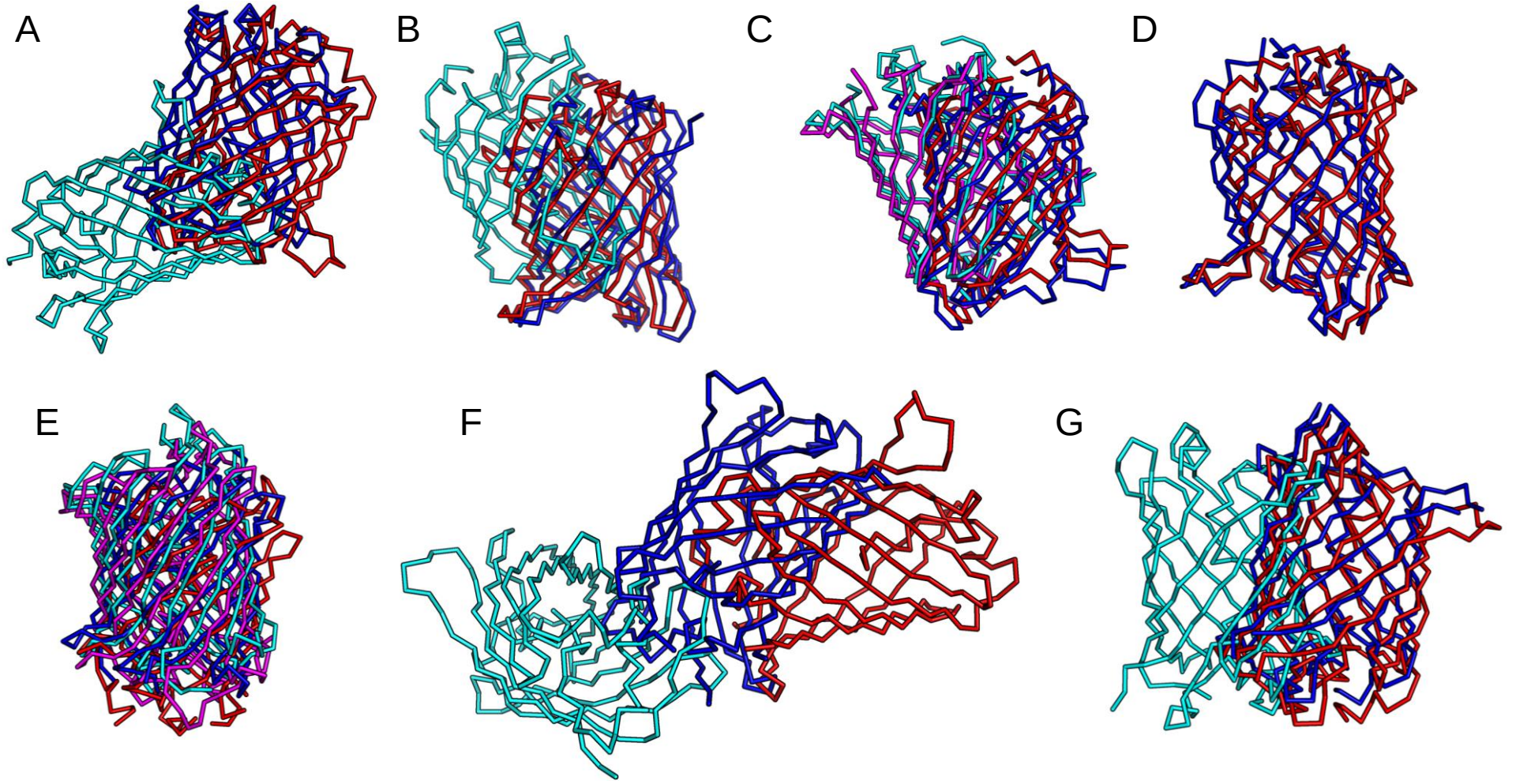


Figure 5

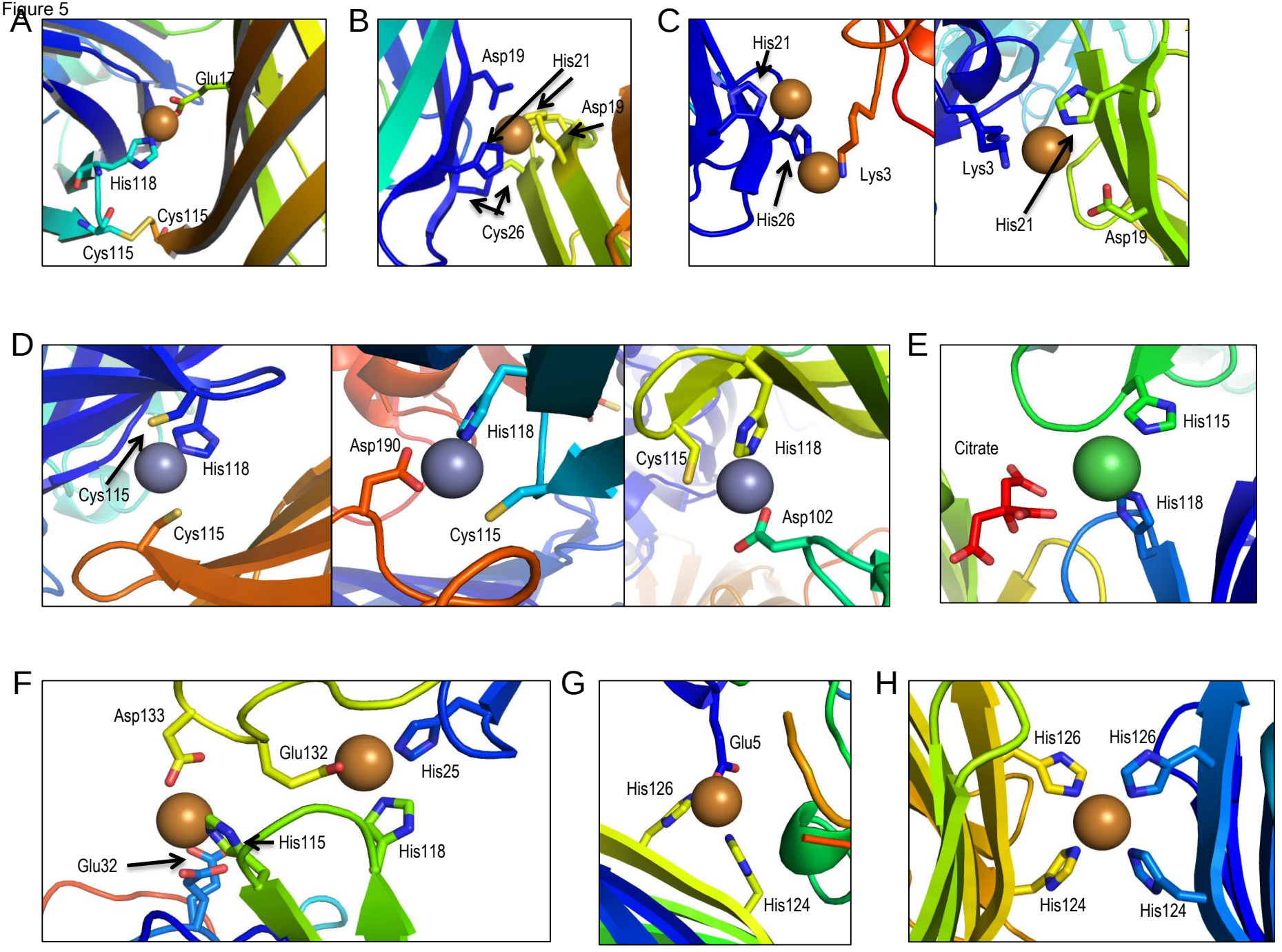
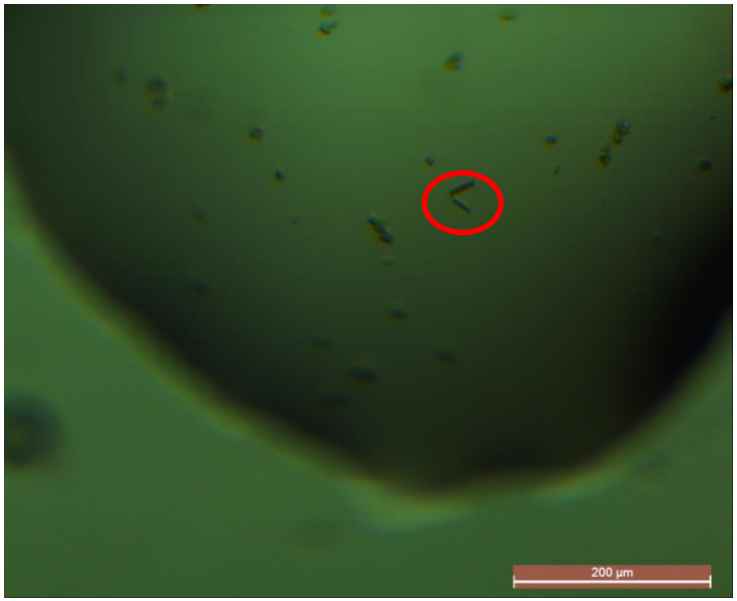

A
Glu17
His118
Cys115
Cys115
B
Asp19
His21
Asp19
Cys26
C
His21
Lys3
His26
Lys3
His21
Asp19
D
His118
Cys115
Cys115
Asp190
His118
Cys115
Cys115
His118
Asp102
E
His115
Citrate
His118
F
Asp133
Glu132
His25
His115
His118
Glu32
G
Glu5
His126
His124
H
His126
His126
His124
His124

Figure 6

A



B

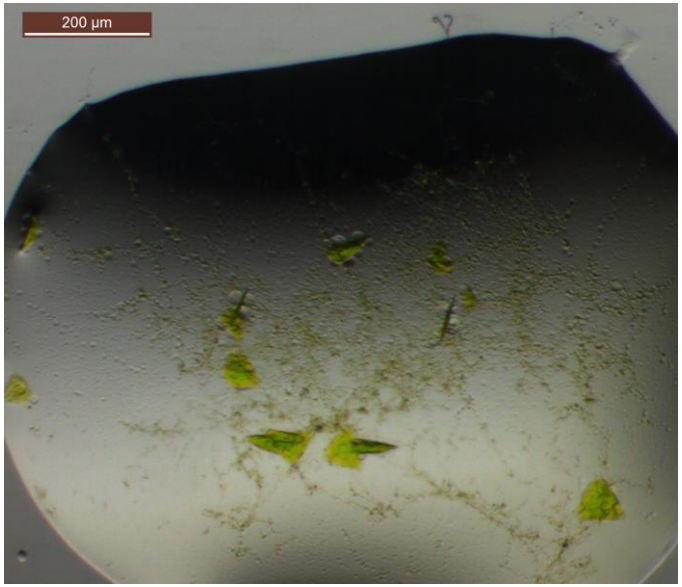
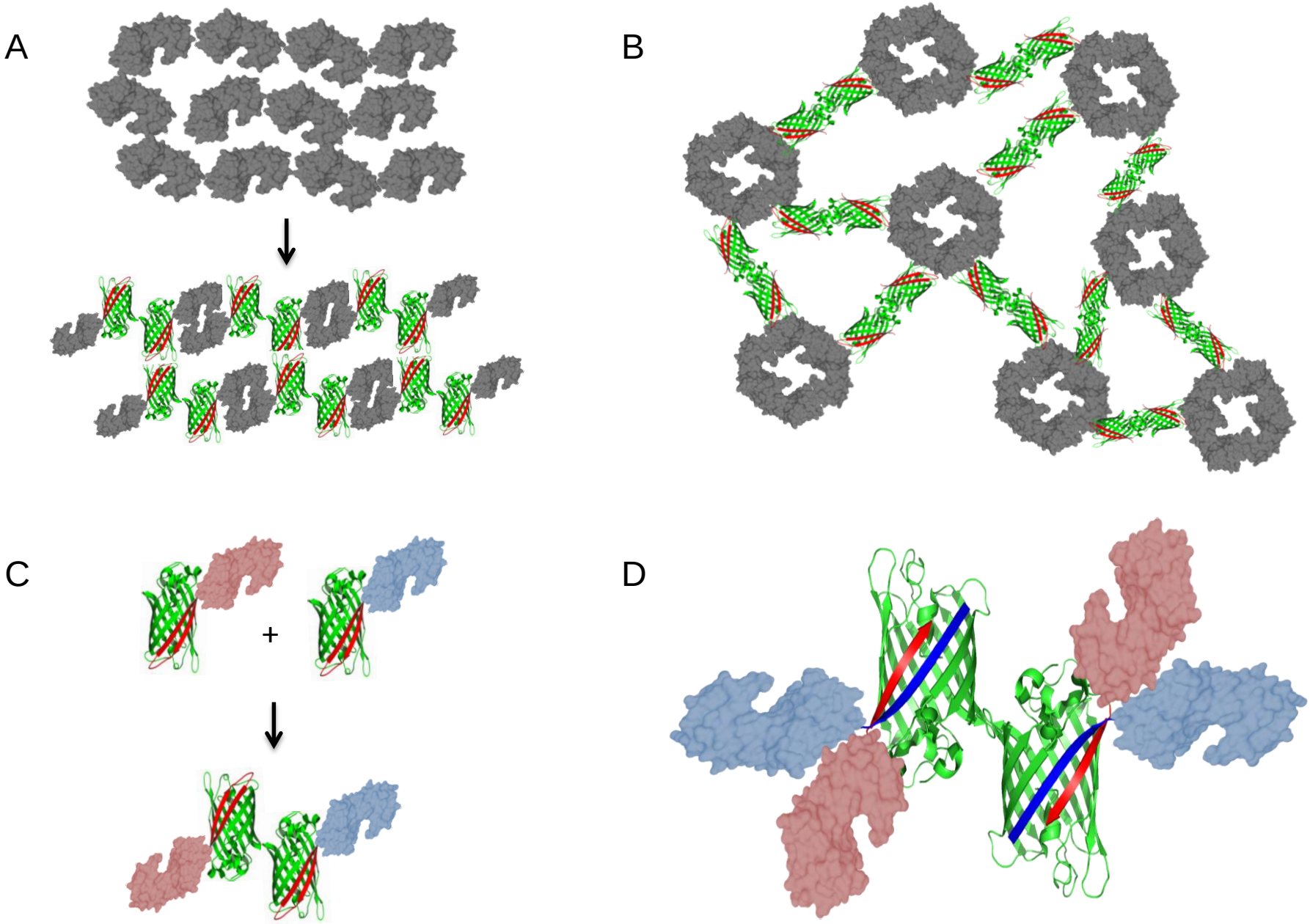


Figure 7



Supplemental Table S1. Data collection and refinement statistics, Related to Table 1.

PDB	4W69	4W6A	4W6B	4W6C	4W6D	4W6F	4W6G	4W6H	4W6I	4W6J	4W6K	4W6L	4W6M	4W6N	4W6O	4W6P	4W6R	4W6S	4W6T	4W6U	4W72	4W73	4W74	4W7X	4W75	4W76	4W77	4W7A	4W7C	4W7D	4W7E	4W7F	4W7R
Wavelength (A)	1.0717	0.9789	0.97918	0.9789	0.9793	0.9792	0.9793	1.0717	1.0717	0.9793	1.0717	1.0717	0.9793	0.9537	0.9793	0.9793	0.9792	0.9789	0.9795	0.9795	0.9795	0.9789	0.9795	0.9789	1.0717	0.9792	0.9789	0.9792	0.9795	0.9792	0.9792	0.9789	0.9789
Resolution range (A)	94.58 - 3.975 (4.117 - 3.975)	77.02 - 2.991 (3.098 - 2.991)	44.69 - 1.895 (1.963 - 1.895)	71.3 - 2.492 (2.581 - 2.492)	87.16 - 3.447 (3.57 - 3.447)	84.34 - 2.701 (2.798 - 2.701)	69.09 - 3.024 (3.132 - 3.024)	82.72 - 1.953 (2.023 - 1.953)	53.59 - 2.625 (2.719 - 2.625)	98.5 - 1.702 (1.763 - 1.702)	75.46 - 2.877 (2.98 - 2.877)	76.67 - 2.45 (2.538 - 2.45)	73.83 - 2.793 (2.893 - 2.794)	88.89 - 3.375 (3.496 - 3.375)	67.51 - 2.6 (2.693 - 2.6)	79.58 - 3.085 (3.195 - 3.085)	89.88 - 3.471 (3.595 - 3.471)	68.28 - 3.1 (3.211 - 3.1)	74.46 - 1.604 (1.661 - 1.604)	82.99 - 2.278 (2.36 - 2.278)	57.12 - 1.996 (2.067 - 1.996)	52.18 - 2.787 (2.887 - 2.787)	88.27 - 2.099 (2.174 - 2.099)	66.77 - 2.8 (2.9 - 2.8)	69.13 - 3.47 (3.597 - 3.473)	60.5 - 2.345 (2.429 - 2.345)	60.79 - 3.1 (3.211 - 3.1)	96.28 - 3.603 (3.731 - 3.603)	96.15 - 2.5 (2.59 - 2.5)	66.57 - 1.799 (1.863 - 1.799)	67.92 - 2.592 (2.685 - 2.592)	48.76 - 2.9 (3.004 - 2.9)	92.07 - 1.799 (1.863 - 1.799)
Space group	P 43 21 2	P 32 2 1	P 21 21 21	P 21 21 21	P 32 2 1	P 32 2 1	P 61	P 65	P 21 21 21	P 31 2 1	P 41 21 2	I 41 2 2	P 63	C 1 2 1	P 64 2 2	P 21 21 21	P 1	P 43 21 2	P 43 21 2	P 21 21 21	P 21 21 21	P 21 21 21	P 1 21 1	P 1 21 1	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	C 1 2 1	P 21 21 21	P 41 21 2	C 2 2 21	P 1 21 1
Unit cell	133.76 133.76 88.92 90 90 90	88.93 88.93 135.76 90 90 120	50.151 90.356 102.83 90 90 90	51.33 88.37 120.69 90 90 90	123.11 123.11 151.32 90 90 120	121.98 121.98 140.09 90 90 120	93.38 93.38 132.97 90 90 120	95.52 95.52 132.5 90 90 120	57.35 67.6 150.58 90 90 90	113.74 113.74 82.46 90 90 120	106.72 106.72 97.45 90 90 90	108.43 108.43 101.47 90 90 90	170.5 170.5 79.57 90 90 120	181.21 102.68 84.13 90 101.44 90	77.95 77.95 178.88 90 90 120	86.05 117.86 209.1 90 90 90	92.42 92.56 124.53 94.94 96.17 102.25	91.51 91.51 205.11 90 90 90	105.3 105.3 69.61 90 90 90	47.64 116.58 165.98 90 90 90	72.54 74.4 89.15 90 90 90	69.74 70.58 77.5 90 90 90	67.44 119.79 130.57 90 89.99 90	66.84 70.47 116.78 90 92.56 90	59.86 83.93 121.88 90 90 90	64.16 66.84 121 90 90 90	62.12 68.29 121.58 90 90 90	120.83 121.33 192.56 90 90 90	205.49 69.42 102.81 90 110.73 90	56.42 82.05 113.87 90 90 90	96.05 96.05 69.96 90 90 90	68.2 69.75 82.57 90 90 90	62.67 87.19 92.07 90 90.01 90
Total reflections	110307 (10414)	252345 (23254)	251856 (13789)	81636 (7060)	89887 (8957)	335541 (32238)	66094 (6525)	508993 (47571)	88235 (6817)	339028 (33029)	167091 (13967)	148335 (14256)	63420 (1852)	70644 (6362)	160637 (16471)	261903 (23019)	87921 (8257)	418083 (42445)	653315 (63113)	279406 (25897)	238183 (20955)	62325 (4525)	411267 (39095)	183888 (18230)	104617 (8741)	145381 (13147)	63735 (6380)	220621 (21172)	321515 (31440)	326510 (29374)	134951 (13211)	23617 (1523)	306630 (30073)
Unique reflections	7344 (715)	12990 (1219)	37298 (3260)	19311 (1713)	17869 (1730)	33538 (3281)	12859 (1284)	49488 (4865)	17334 (1613)	67258 (6633)	13200 (1210)	11419 (1106)	32658 (378)	20957 (1964)	10518 (991)	39143 (3478)	46082 (4386)	16549 (1612)	51580 (4985)	42930 (4011)	33399 (3272)	9864 (848)	119482 (11566)	26887 (2649)	8254 (755)	22306 (2125)	9841 (949)	33244 (3177)	46757 (4596)	49634 (4736)	10580 (999)	4558 (378)	90790 (8888)
Multiplicity	15.0 (14.6)	19.4 (19.1)	6.8 (4.2)	4.2 (4.1)	5.0 (5.2)	10.0 (9.8)	5.1 (5.1)	10.3 (9.8)	5.0 (4.2)	5.0 (5.0)	12.7 (11.5)	13.0 (12.9)	5.7 (4.9)	3.4 (3.2)	15.3 (16.6)	6.7 (6.6)	1.9 (1.9)	25.3 (26.3)	12.7 (12.7)	6.5 (6.5)	7.1 (6.4)	6.3 (5.3)	3.4 (3.4)	6.8 (6.9)	12.7 (11.6)	6.5 (6.2)	6.5 (6.7)	6.6 (6.7)	6.9 (6.8)	6.6 (6.2)	12.8 (13.2)	5.4 (4.0)	3.4 (3.4)
Completeness (%)	99.90 (99.31)	99.52 (95.08)	98.40 (87.52)	97.33 (88.79)	99.26 (98.69)	99.95 (99.64)	99.74 (99.46)	99.80 (98.06)	95.43 (82.54)	99.47 (98.82)	99.25 (93.51)	99.83 (98.57)	99.15 (95.61)	97.89 (91.99)	99.94 (99.70)	97.98 (88.91)	89.44 (85.24)	99.95 (99.94)	99.78 (97.98)	99.28 (94.55)	99.74 (98.55)	98.67 (87.69)	98.81 (95.89)	99.66 (99.62)	98.78 (94.83)	99.23 (96.33)	99.87 (99.79)	99.40 (95.49)	99.06 (97.93)	99.57 (96.26)	99.68 (97.18)	99.52 (99.55)	98.83 (97.73)
Mean I/sigma(I)	16.94 (2.10)	20.34 (2.49)	12.35 (4.49)	7.28 (1.91)	11.91 (1.67)	5.64 (1.97)	16.73 (2.42)	16.42 (2.95)	5.91 (1.02)	16.58 (2.07)	22.07 (2.60)	23.95 (3.39)	4.81 (3.81)	6.64 (1.48)	25.97 (3.43)	10.85 (2.19)	5.70 (1.38)	18.03 (3.63)	17.81 (1.58)	14.20 (1.91)	16.59 (2.40)	10.07 (1.72)	8.06 (1.46)	8.19 (1.57)	13.64 (1.74)	10.12 (1.82)	9.86 (2.58)	13.95 (2.28)	15.63 (1.87)	8.80 (0.97)	19.81 (2.01)	7.98 (4.02)	6.36 (1.20)
Wilson B-factor	162.13	72.26	20.64	60.27	112.78	64.39	95.48	25.44	54.36	25.17	91.58	61.79	71.33	84.61	68.93	70.73	101.51	76.17	26.53	44.65	41.24	73.93	32.25	57.13	125.59	47.91	69.36	112.51	63.16	25.91	58.95	82.37	23.54
R-merge	0.144 (1.663)	0.1626 (1.385)	0.1043 (0.376)	0.1036 (0.6601)	0.1177 (1.023)	0.3045 (0.5784)	0.06416 (0.7286)	0.1088 (0.8194)	0.2459 (1.047)	0.05086 (0.7019)	0.08328 (1.265)	0.06428 (0.8398)	0.7659 (0.7891)	0.2049 (0.7778)	0.07197 (0.9344)	0.1491 (0.8467)	0.09489 (0.4669)	0.257 (1.844)	0.07155 (1.129)	0.1054 (0.9065)	0.05998 (0.8213)	0.1087 (0.8508)	0.1043 (0.7747)	0.1897 (1.297)	0.1496 (1.419)	0.1097 (1.015)	0.1514 (0.7477)	0.1225 (0.7986)	0.0733 (1.03)	0.138 (1.868)	0.1132 (1.67)	0.1781 (0.4051)	0.1222 (1.021)
R-meas	0.1491	0.1669	0.1131	0.1182	0.1317	0.3213	0.07156	0.1145	0.2731	0.05696	0.08686	0.06706	0.8229	0.2439	0.07454	0.1618	0.1341	0.2623	0.07467	0.115	0.06477	0.1187	0.1235	0.2054	0.1561	0.1195	0.1648	0.1329	0.07941	0.15	0.118	0.1958	0.1454
CC1/2	0.999 (0.714)	0.999 (0.767)	0.995 (0.871)	0.989 (0.881)	0.998 (0.607)	0.955 (0.871)	0.998 (0.853)	0.998 (0.811)	0.983 (0.853)	0.999 (0.782)	0.999 (0.811)	0.999 (0.979)	0.683 (0.49)	0.992 (0.746)	0.999 (0.938)	0.994 (0.76)	0.989 (0.773)	0.998 (0.947)	0.999 (0.821)	0.997 (0.739)	0.999 (0.895)	0.996 (0.7)	0.996 (0.747)	0.991 (0.616)	0.999 (0.944)	0.998 (0.924)	0.996 (0.814)	0.998 (0.804)	0.999 (0.917)	0.997 (0.451)	0.999 (0.792)	0.98 (0.826)	0.994 (0.71)
CC*	1 (0.913)	1 (0.932)	0.999 (0.965)	0.997 (0.968)	1 (0.869)	0.988 (0.965)	1 (0.959)	1 (0.947)	0.996 (0.959)	1 (0.937)	1 (0.946)	1 (0.995)	0.901 (0.811)	0.998 (0.924)	1 (0.984)	0.999 (0.929)	0.997 (0.934)	1 (0.986)	1 (0.95)	0.999 (0.922)	1 (0.972)	0.999 (0.908)	0.999 (0.925)	0.998 (0.873)	1 (0.985)	1 (0.98)	0.999 (0.947)	0.999 (0.944)	1 (0.978)	0.999 (0.789)	1 (0.94)	0.995 (0.951)	0.998 (0.911)
R-work	0.3070 (0.4575)	0.1912 (0.3043)	0.1670 (0.2013)	0.2480 (0.4428)	0.2360 (0.3634)	0.2040 (0.2584)	0.2480 (0.3844)	0.1660 (0.2002)	0.2680 (0.5039)	0.1890 (0.2588)	0.2491 (0.3511)	0.2521 (0.3755)	0.2610 (0.3886)	0.3162 (0.3996)	0.2620 (0.3484)	0.2323 (0.3154)	0.3070 (0.4087)	0.2232 (0.2836)	0.1800 (0.2599)	0.2100 (0.2768)	0.1900 (0.2920)	0.2210 (0.3661)	0.2120 (0.3034)	0.2170 (0.3173)	0.3015 (0.4443)	0.2330 (0.4241)	0.2169 (0.2600)	0.2780 (0.3358)	0.2265 (0.4082)	0.1790 (0.3174)	0.2070 (0.3534)	0.2640 (0.4037)	0.2230 (0.3765)
R-free	0.3350 (0.3978)	0.2398 (0.3891)	0.2020 (0.2803)	0.2761 (0.4453)	0.2670 (0.3179)	0.2380 (0.2898)	0.2700 (0.3607)	0.1900 (0.2288)	0.3160 (0.6061)	0.2120 (0.2891)	0.2941 (0.3949)	0.2780 (0.4499)	0.2850 (0.3936)	0.3626 (0.4687)	0.3110 (0.4147)	0.2789 (0.3628)	0.3570 (0.4314)	0.2760 (0.3716)	0.2070 (0.2890)	0.2500 (0.3192)	0.2350 (0.3136)	0.2970 (0.4641)	0.2350 (0.3298)	0.2690 (0.3857)	0.3446 (0.3766)	0.2880 (0.4588)	0.2908 (0.3775)	0.3018 (0.3366)	0.2537 (0.4281)	0.2210 (0.3417)	0.2620 (0.4414)	0.3320 (0.4394)	0.2530 (0.4255)
Number of non-hydrogen atoms	3458	3574	3867	3553	3550	3604	3505	3884	3558	3925	3037	1635	6752	10419	1662	12960	25002	3538	2074	7317	3817	3519	14583	7089	3181	3639	3474	7085	7028	4014	1820	1726	7625
macromolecules	3414	3530	3599	3509	3505	3539	3461	3635	3514	3623	2993	1613	6652	10331	1637	12828	25002	3442	1846	7083	3570	3469	14200	7001	3180	3588	3473	6994	6938	3603	1766	1703	7166
ligands	44	44	47	44	45	65	44	44	44	96	44	22	100	88	22	132	0	96	69	111	45	50	182	88	1	45	1	91	90	103	28	23	146
water	0	0	221	0	0	0	0	205	0	206	0	0	0	0	3	0	0	0	159	123	202	0	201	0	0	6	0	0	0	309	26	0	313
Protein residues	434	446	454	443	445	445	437	457	443	451	378	205	842	1306	208	1618	3133	436	227	890	450	437	1793	882	396	452	432	881	873	224	222	215	677
RMS(bonds)	0.011	0.018	0.009	0.01	0.01	0.011	0.011	0.011	0.014	0.011	0.011	0.01	0.011	0.009	0.013	0.009	0.008	0.012	0.019	0.012	0.011	0.013	0.013	0.009	0.004	0.011	0.01	0.011	0.013	0.01	0.012	0.01	0.012
RMS(angles)	1.29	1.6	1.08	1.29	1.69	1.37	1.37	1.21	1.3	1.21	0.95	1.23	1.2	1.49	1.85	1.42	1.76	1.55	1.7	1.36	1.24	1.42	1.51	0.93	0.85	1.28	1.4	1.22	1.33	1.26	1.46	1.37	1.27
Ramachandran favored (%)	97	97	98	96	98	97	97	98	97	99	97	96	99	95	98	97	96	95	99	98	98	97	98	97	98	99	94	96	99	97	97	97	99
Ramachandran outliers (%)	0	0.23	0	0.23	0.7	0.23	0.48	0	0	0	0.29	0	0.12	0.08	0	0.13	0.44	0.47	0	0.12	0	0	0	0.23	0	0	0.24	0.35	0	0	0	0	0
Clashscore	21.24	14.17	1.39	12.18	17.52	6.2	21.22	1.52	5.41	4.88	9.3	8.07	13.97	30.03	20.8	12.63	19.98	21.62	6.31	4.52	3.23	13.73	7.24	11.33	4.14	9.07	12.4	7.98	15.05	3.96	7.08	10.93	5.96
Average B-factor	191..00	64	25	63	124.1	68.6	171.7	26.9	56.1	33.9	94.8	93.1	91.1	33.4	106.4	80.1	113.2	84.2	32	47.4	46.4	67.8	39.1	50.7	161.6	64.7	65.2	117.2	100.1	21.3	52.7	74.7	33.5
macromolecules	191.5	64.1	24.9	63.2	124.3	68.4	171.7	26.8	56.3	33.5	35.1	93.3	91.4	99.6	106.6	80.2	113.2	84	31	47.7	46.2	67.9	39.1	50.8	161.6	64.8	65.2	118.2	100.3	30.4	52.8	75	33.2
ligands	147.9	54.6	18.5	50	107.6	77.1	175.5	20.1	42.6	38.9	78.3	79.2	72	72.8	97.2	70.6	--	91	42	36.4	42.4	55.9	36.3	43.3	196.4	61.1	64.9	42.2	83.9	39.9	53.4	51	37.1
solvent	--	--	27.8	--	--	--	--	30.7	--	37.1	--	--	--	--	72	--	--	--	39.7	43.8	50.5	--	37.1	---	--	56.8	--	--	--	39	46.2	--	38

Supplemental Table S2. Disulfide bond dihedral angle energy server output, Related to Table 2.

PDB	Cys1	Chi1(X1)	Chi2(X2)	Chi3(X3)	Bond Distance (Å)	Chi2'(X2')	Chi1'(X1')	Cys2	Disulfide Strain Energy (kJ/mol)
4W69	157:A:	-154.01	30.11	-178.47	2.04	-89	-77.54	157:B:	47.299248
4W6A	157:A:	-177.00	73.24	38.39	2.31	73.28	-177.02	157:B:	21.57205
4W6B	26:A:	-72.82	86.33	78.75	2.05	98.94	78.42	26:B:	17.093836
4W6C	26:A:	75.31	78.48	106.19	2.07	75.11	67.4	26:B:	12.889212
4W6D	26:A:	55.17	-177.57	143.87	2.05	93.86	-48.99	26:B:	29.07711
4W6F	26:A:	-72.5	-171.2	-78.52	2.04	132.24	-75.52	26:B:	14.561133
4W6G	190:A:	-70.19	-42.67	-92.93	2.17	-86.15	-59.9	190:B:	9.081745
4W6H	190:A:	-31.17	-105.20	-87.99	2.05	-35.15	-86.13	190:B:	27.061176
4W6I	190:A:	-64.99	142.91	100.01	2.03	151.48	-63.81	190:B:	14.652457
4W6J	117:A:	-54.02	-55.14	-92.71	2.02	-109.94	-61.84	117:B:	11.304162
4W6K	117:A:	-75.04	-66.27	-127.5	2.23	-18.1	-84.4	117:B:	30.940155
4W6L	117:A:	-41.90	-81.68	-113.18	1.93	-81.68	-41.90	117:B:	21.203020
4W6M	117:A:	-83.09	-48.39	-114.58	2.03	-73.8	-82.88	117:C:	22.50606
4W6M	117:B:	-163.74	46.93	159.96	2.03	-94.13	-83.79	117:D:	41.670086
4W6N	117:A:	-42.79	91.35	-150.87	2.05	-85.41	55.02	117:D:	35.854893
4W6N	117:B:	-66.27	82.17	106.38	2.03	145.25	-65.11	117:F:	15.203947
4W6O	117:A:	79.69	56.05	-110.65	2.03	56.05	-79.69	117:B:	16.677874
4W6P	102:A:	-82.1	-7.53	-124.45	2.03	47.95	-89.95	102:B:	36.520576
4W6P	102:C:	-87.56	45.31	-121.68	2.02	-5.68	-80.3	102:D:	34.042976
4W6P	102:E:	-78.91	-28.57	-128	2.01	87.91	-152.25	102:H:	35.283493
4W6P	102:F:	-65.00	-37.24	-62.86	2.02	5.98	-64.95	102:G:	17.474335
4W6R	102:A:	-83.29	-8.38	-87.58	2.11	-124.65	70.4	102:N:	25.146637
4W6R	102:B:	-69.22	16.41	-114.52	1.96	15.91	-70.74	102:I:	26.10725
4W6R	102:C:	-66.76	15.51	-110.58	2.02	21.99	-69.48	102:D:	22.807249
4W6R	102:E:	-66.97	26.66	-124.7	2.18	27.85	-78.86	102:J:	28.300119
4W6R	102:F:	-68.49	-16.22	-164.19	1.99	93.94	-99.35	102:O:	53.013672
4W6R	102:G:	-27.15	-88.45	135.51	1.99	-179.47	-69.5	102:P:	33.607735
4W6R	102:H:	-55.13	-32.47	-127.85	1.98	111.91	-158.85	102:M:	32.367477
4W6R	102:K:	-67.4	18.31	-113.65	2.02	17.26	-67.04	102:L:	24.121847
4W6S	126:A:	-80.49	-84.83	-70.22	2.04	-79.74	-167.69	126:B:	14.860621
4W7X	115:A:	-60.4	-30.1	-96.94	2.04	-132.62	161.69	115:B:	19.031637
4W7X	115:C:	-72.82	-80.96	105.38	2.23	106.57	55.01	115:D:	18.124321
4W72	115:A:	174.93	-130.88	-90.68	2.05	-58.13	-62.57	115:B:	10.6925
4W73	115:A:	-176.42	-137.86	-86.8	2.14	-53.57	-68.97	115:B:	10.131399
4W75	26:A:	-74.50	110.78	152.70	2.05	-121.01	-26.35	26:B:	53.769878
4W7C	26:A:	66.77	107.3	-162.69	2.01	-102.12	-171.58	26:B:	43.09621
4W7C	26:C:	78.93	71.48	80.82	1.99	92.15	74.27	26:D:	13.494884
4W76	26:A:	86.6	82.03	91.34	1.92	72	75.12	26:B:	15.361248
4W77	26:A:	-161.27	-84.86	-39.65	2.04	-98.47	-157.73	26:B:	36.480595
4W7A	26:A:	-150.96	-114.06	-177.41	2.03	85.41	165.7	26:B:	50.774982
4W7A	26:C:	174.67	-96.21	-159.9	2.03	124.31	81.24	26:D:	45.845509

Values of the disulfide bond dihedral angles and calculated bond energies of each disulfide dimer as determined by the Disulfide Bond Dihedral Angle Energy Server (http://services.mbi.ucla.edu/disulfide/).

Supplemental Table S3. Angular variations between dimers in pairwise comparisons, Related to Table 2.

D102C	4W6R_BI	4W6R_CD	4W6R_EJ	4W6R_FO	4W6R_GP	4W6R_HM	4W6R_KL	4W6P_AB	4W6P_CD	4W6P_EH	4W6P_FG
4W6R_AN	4	7.3	2.7	3.5	3.4	6	7.7	25.7	24	25.2	28.5
4W6R_BI		5.7	3.9	1.7	2.1	2.3	5.9	23.8	21.8	23.4	27.2
4W6R_CD			5.1	4.8	7	4.6	1.2	28.6	26.3	28.1	32
4W6R_EJ				3.2	4.6	5.2	5.3	27.4	25.5	26.9	30.5
4W6R_FO					2.2	2.7	5.3	24.9	22.9	24.4	28.1
4W6R_GP						3.9	7.5	23	21.1	22.4	26
4W6R_HM							4.8	24.1	21.8	23.7	27.7
4W6R_KL								28.8	26.5	28.4	32.4
4W6P_AB									3	2.6	6
4W6P_CD										3.7	8.3
4W6P_EH											4.7

D190C	4W6H	4W6G
4W6I	41.4	38.1
4W6H		6.3

K26C	4W6B	4W6F	4W6D
4W6C	136.1	33.3	20.7
4W6B		137.1	140.4
4W6F			19

Q157C	4W6A_A	4W6A_B
4W69	49	129
4W6A_A		95.8

D117C	4W6N_AD	4W6N_BF	4W6N_CE	4W6M_AC	4W6M_BD	4W6L	4W6K	4W6J
4W6O	32.6	34.2	34	12	11.1	7	16.4	25.7
4W6N_AD		2	1.6	33.4	31.9	33.2	20.2	9.5
4W6N_BF			1	34.8	33.3	34.7	21.6	10.8
4W6N_CE				34.7	33.4	34.6	21.7	10.3
4W6M_AC					4.1	5.3	13.7	29
4W6M_BD						4.8	11.8	27.7
4W6L							14.1	27.8
4W6K								17.6

E115C	4W7X_AB	4W7X_CD	4W72
4W73	8.7	10.3	12.3
4W7X_AB		6.7	8.6
4W7X_CD			4.6

21hc	4W7A_AB	4W7A_CD	4W76	4W7C_AB	4W7C_CD	4W75
4W77	6.9	6.1	4	6.9	5.4	30.7
4W7A_AB		8.6	5.5	5.2	2.5	24.2
4W7A_CD			3.3	6.7	6.3	32.1
4W76				4.6	3.3	29.5
4W7C_AB					4.9	27.8
4W7C_CD						26.3

The values shown are in degrees. The structures being compared are designated by their PDB code followed by the chain identifiers for the two subunits in a dimeric arrangement.

Supplemental Table S4. RMS coordinate deviations between dimers arising from rotational angle variations, Related to Table 2.

D102C	4W6R_BI	4W6R_CD	4W6R_EJ	4W6R_FO	4W6R_GP	4W6R_HM	4W6R_KL	4W6P_AB	4W6P_CD	4W6P_EH	4W6P_FG
4W6R_AN	1.7	2.2	0.9	1.6	0.9	1.7	2.7	8.5	8.3	6.8	9
4W6R_BI		1.9	1.3	0.6	1.1	0.8	1.9	8.1	7.8	6.8	9
4W6R_CD			1.6	2.1	2	1.5	0.9	9.8	9.5	8.3	10.5
4W6R_EJ				1.2	0.9	1.2	2	8.7	8.4	7.2	9.3
4W6R_FO					1	1	2.1	7.9	7.6	6.6	8.8
4W6R_GP						1.2	2.4	8.1	7.9	6.5	8.7
4W6R_HM							1.5	8.5	8.2	7.1	9.3
4W6R_KL								9.7	9.4	8.4	10.7
4W6P_AB									0.6	2.7	2.8
4W6P_CD										2.7	3.1
4W6P_EH											2.5

D190C	4W6H	4W6G
4W6I	20	19
4W6H		3.9

K26C	4W6B	4W6F	4W6D
4W6C	36.4	8.2	6.2
4W6B		35.3	36.74
4W6F			7.7

Q157C	4W6A_A	4W6A_B
4W69	19.6	50.9
4W6A_A		38

D117C	4W6N_AD	4W6N_BF	4W6N_CE	4W6M_AC	4W6M_BD	4W6L	4W6K	4W6J
4W6O	10.9	12	11.7	2.6	2.7	1.6	6	10.1
4W6N_AD		1.5	1.3	11.7	9.9	11	5.8	3.8
4W6N_BF			0.4	12.8	10.9	12	6.6	3.5
4W6N_CE				12.5	10.7	11.8	6.4	3.4
4W6M_AC					2.5	1.7	6.5	11.2
4W6M_BD						1.9	4.4	9.3
4W6L							5.7	10.3
4W6K								5.3

E115C	4W7X_AB	4W7X_CD	4W72
4W73	2.5	3.4	3.8
4W7X_AB		2.7	2.7
4W7X_CD			1.5

21hc	4W7A_AB	4W7A_CD	4W76	4W7C_AB	4W7C_CD	4W75
4W77	2.6	2.6	1.8	2.5	2.2	16.3
4W7A_AB		3.7	2.1	1.3	0.8	14
4W7A_CD			1.7	3.1	3.1	17.3
4W76				1.6	1.5	15.9
4W7C_AB					1.2	14.7
4W7C_CD						14.6

The values shown are in Angstroms. The RMSD values represent the deviation of the Cα alignments used for the pairwise comparisons presented in Table S3.

Supplemental Table S5. Cloning primers, Related to Methods.

Primer Name	Primer Sequence	
GFP.For.	5’-ggaattacatatgaggaaaggagaagaac-3'	
GFP.Rev.	5’-ttttttaagcttctattaatggtgatggtgatgatgtgtaatcccagcagcagttac-3'	
C48A.For.New.	5’-gccactactggaaaactacctgttcc-3'	
C48A.Rev.New.	5’-aataaatttaaggctgagttttccg-3'	
C70A	5'-tctgacctatggtgttcaagccttttcccgttatccggat-3'	
C70A_antisense	5'-atccggataacgggaaaaggcttgaacaccataggtcaga-3'	
D21H	5'-caattcttattgaattagatggtcatgttaatgggcactgctttttt-3'	C48A/K26C/D21H
D21H_antisense	5'-aaaaaagcagtgcccattaacatgaccatctaattcaataagaattg-3'	
D102C	5'-ttatgtacaggaacgcactatatatttcaaatgtgacgggacctacaag-3'	(C48A/C70A/D102C)
D102C_antisense	5'-cttgtaggtcccgtcacatttgaaatatatagtgcgttcctgtacataa-3'	
D117C	5'-tgctgaagtcaagtttgaaggttgtacccttgttaatcgtatcgag-3'	(C48A/D117C/C70A)
D117C_antisense	5'-ctcgatacgattaacaagggtacaaccttcaaacttgacttcagca-3'	
Q157C	5'-cacaaagtatacatcacggcagacaaatgcaataatggaatcaaagctaacttcaca-3'	C48A/C70A/Q157C)
Q157C_antisense	5'-tgtgaagttagctttgattccattattgcatttgtctgccgtgatgtatactttgtg-3'	
K26C	5'-gatggtgatgttaatgggcactgctttttgtccgtggagagggt-3'	(C48A/K26C/C70A)
K26C_antisense	5'-accctctccacggacaaaaaagcagtgcccattaacatcaccatc-3'	
D190C	5'-aacaaaatactccaattggctgtggccctgtcctttaccag-3'	(C48A/D190C/C70A)
D190C_antisense	5'-ctggtaaaaggacagggccacagccaattggagtattttgtt-3'	
E124H.K126H.For.	5'-agtttgaaggtgataccttgttaatcgtatccatttacatggtattgattttaaagaagatggaaacattc-3'	C48A/C70A/E124H/K126H
E124H.K126H.Rev.	5'-gaatgtttccatcttctttaaaatcaataccatgtaaatggatacgattaacaagggtatcaccttcaaact-3'	
E115H.T118H.For.	5'-acaagacgcgtgctgaagtcaagtttcatggtgatcaccttgttaatcgtatcg-	C48A/C70A/E115H/T118H
E115H.T118H.Rev.	5'-cgatacgattaacaaggtgatcaccatgaaacttgacttcagcacgcgtcttgt-3	
H115C.For.	5'-caagacgcgtgctgaagtcaagttttgtggtgatcacctt-3'	Used with above primers to make: C48A/C70A/E115C/T118H
H115C.Rev.	5'-aaggtgatcaccacaaaacttgacttcagcacgcgtcttg-3'	
GFP.pMA507-star.For.	5’- aaaacctgtacttccagggcatgaggaaaggagaagaacttttcac-3'	
GFP.pMA507-star.Rev.	5’-aacgagttaattaagtcgcgttatgtaatcccagcagcagttacatac-3'	
PIPE.Vec.For.	5’-cgcgacttaattaactcgtttaaacggtctccagc-3'	
PIPE.Vec.Rev.	5’-ctggaagtacaggttttcgtgatgatgatgatgatg-3'	

Sequences of the primers used for cloning the suite of GFP mutants

Supplemental Table S6. Crystallization and cryo-protectant conditions, Related to Table 1.

PDB	Crystal Condition	Cryo Protectant
4W69	0.4M MgFormate, 0.1M Acetate pH 4.6, 2%w/v benzamidine	25%v/v Glycerol
4W6A	2.0M NaFormate, 0.1M NaAcetate pH 4.6	25%v/v Glycerol
4W6B	14%w/v PEG-4000, 0.2M MgCl2, 0.1M Tris pH 8.5	25%v/v Glycerol
4W6C	35%v/v MPD, 0.1M Imidazole pH 8.0, 0.2M MgCl2	--
4W6D	1.5M MgSO4, 0.5%w/v Glycerol, 0.1M MES pH 6.75	30%v/v Glycerol
4W6F	10%v/v 2-propanol, 0.1M Imidazole pH 8.0	25%v/v Glycerol
4W6G	20%v/v 1,4-Butanediol, 0.1M Acetate pH 4.5	20%v/v Glycerol
4W6H	0.1M SPG Buffer pH8.0, 25%w/v PEG-1500	--
4W6I	1.4M MgSO4, 0.1M BTP pH 7.4	25%v/v Glycerol
4W6J	35% MPD, 0.1M NaAcetate pH 4.5	--
4W6K	0.5M KSCN, 0.1M NaAcetate pH 4.6	25%v/v Glycerol
4W6L	1.5M NaNO3, 0.1M NaAcetate pH 5.0	25%v/v Glycerol
4W6M	10%w/v PEG3350, 0.1M NaAcetate pH 4.6, 0.2M NaCl	25%v/v Glycerol
4W6N	10%w/v PEG6000, 0.1M Hepes pH 6.5	20%v/v Glycerol
4W6O	20%w/v PEG6000, 0.1M Bicine pH 8.5	20%v/v Glycerol
4W6P	1.3M NaNO3, 0.1M NaAcetate pH 5.0	25%v/v Glycerol
4W6R	20%w/v PEG3350, 0.2M NaSCN	25%v/v Glycerol
4W6S	40%w/v PEG300, 0.1M Phosphate-citrate pH 4.2	20%v/v Glycerol
4W6T	0.15M Kbr, 30%w/v PEG MME 2000	25%v/v Ethylene glycol
4W6U	0.2M NaCl, 0.1M Phosphate-citrate pH 4.2, 20%w/v PEG8000	25%v/v Ethylene glycol
4W72	20%w/v PEG3000, 0.1M Acetate pH 4.5	25%v/v Glycerol
4W73	20%w/v PEG1000, 0.1M Imidazole pH8.0, 0.2M Ca(OAc)2	--
4W74	17%w/v PEG10000, 0.1M NH4(OAc), 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W7X	1.0 M (NH4)2HPO4, acetate pH 4.5	25%v/v Glycerol
4W75	30%w/v PEG MME 2000, 0.15M KBr	20%v/v Glycerol
4W76	3M NaCl, 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W77	50%v/v PEG200, 0.2M MgCl2, 0.1M NaCacodylate pH 6.5	--
4W7A	3M NaCl, 0.1M Bis-Tris pH 5.5	25%v/v Glycerol
4W7C	30%v/v PEG400, 0.1M Cacodylate pH 6.5, 0.2M Li2SO4	--
4W7D	20%w/v PEG8000, 0.1M CHES pH9.5	25%v/v Glycerol
4W7E	0.1M Imidazole pH 8.0, 10%w/v PEG8000	25%v/v Glycerol
4W7F	20%w/v PEG8000, 0.1M CHES pH9.5	25%v/v Glycerol
4W7R	20%w/v PEG3350, 0.2M Potassium formate	25%v/v Ethylene glycol

Crystallization and cryo-protectant conditions for each structure